CHAPTER 15

OPTICS IN THE NINETEENTH CENTURY

JED Z. BUCHWALD

15.1 The Eighteenth-Century Background

Two fundamental physical images governed speculation in optics, and occasionally even mathematization, from the eighteenth through the nineteenth centuries: namely, the conception of light as a sequence of material particles moving through a void, on the one hand, and the conception of light as a mechanical disturbance in an all-encompassing medium, on the other. The latter image, in a myriad of forms, had by far the greater number of adherents until well into the eighteenth century, and has its roots in René Descartes' comprehensive mechanical system. In 1690 the speculative Cartesian optical medium acquired a novel character when the disturbance it was supposed to carry was bound to geometry by the Dutch polymath Christiaan Huygens.[1] To do so he introduced a physico-mathematical rule, eponymously termed Huygens' Principle, that governed the propagation of the optical disturbances, and according to which each point on the surface of a propagating pulse of light itself constitutes a secondary source, with the overall pulse being the common tangent to all of these secondaries. Huygens thought his disturbances to constitute what we now term longitudinal pulses, which are isolated disturbances that parallel the direction of their propagation. These pulses had no periodic properties, and indeed Huygens' theory was able to deal neither with colours nor with certain curious phenomena that will shortly be critical for us here and that he had himself discovered on passing light through exotic crystals brought from Iceland. Huygens, however, successfully produced a thoroughly geometrical theory, buttressed by careful experiment, for the peculiar double images produced by these Iceland crystals. It is important to note, though, that the computational tools of the day were inadequate to probe the recesses of Huygens' claims, and his construction for double refraction remained controversial until the beginning of the nineteenth century, when, as we shall see, its confirmation in Paris set in motion a significant chain of events.

During the eighteenth century a considerable amount of speculative natural philosophy was produced, but mathematical optics remained for the most part bound to the physical concept that Huygens' system had in fact demoted from physical primacy; namely, the ray of light. The ray itself had long been the foundation of geometrical optics, and in the seventeenth century it had acquired a new physical reality within the first system mentioned above; namely, as marking the track of the particles out of which Isaac Newton built light. This conception of light's structure was extraordinarily influential, inasmuch as it formed an essential part of Newtonian natural philosophy, if not of Newton's mathematical optics. But mathematical optics of any kind was not extensively pursued in new ways during the eighteenth century, and certainly no novel experimental or mathematical results were produced during the period that attracted widespread attention, though significant instrumental developments certainly did occur. Moreover, throughout much of the eighteenth century, in a number of loci the differences between the Newtonian theory and systems based on motion through a medium were not altogether clear-cut, not least because elements of both appear in Newton's own, widely-read *Opticks* (the first edition of which was printed in 1704), in what many readers evidently found to be a confusing amalgam.[2]

Despite (or perhaps because) of the concentration on physical principles during the period, until the last quarter of the eighteenth century very little work in any area of natural philosophy associated with the laboratory was quantitative, and even less attempted to integrate quantitative detail with precise experimental situations whose accuracy could be specified. This was particularly true for investigations of electricity and heat, and it was also true in comparatively obscure areas such as the optics of crystals. This began to change radically in France during the last quarter of the century. Charles Coulomb in electricity and the Lavoisier–Laplace collaboration in heat exemplify this change: here we find a growing concern with quantitative structure coupled to careful, and often elaborate, experiments designed explicitly to concentrate on quantity. Indeed, by the turn of the century in France, work that was not quantitative, and experiments that were not carefully contrived and mathematically analysed, stood little chance of receiving much attention. This new desideratum was best learned by example, and one place to learn it was at the École Polytechnique in Paris. All four of the major French participants in the early years of the optics controversies that reshaped the discipline in fundamental ways—namely, Jean Baptiste Biot, Etienne Louis Malus, François Arago, and Augustin Jean Fresnel—attended the École in the 1790s and early 1800s.

15.2 Ray Optics, the Discovery of Polarization, and the Biot–Arago Controversy

During the first decade of the nineteenth century, interest in optics, and particularly in novel optical experiments, became quite strong in France. Stimulated in part by the English chemist William Hyde Wollaston's apparent confirmation of Huygens' construction for the double refraction of Iceland crystal, Laplace had Malus, in whom he reposed a considerable amount of confidence for work that Malus had already done in constructing a mathematics for systems of light rays, to undertake a thorough experimental investigation of the subject. After producing at Laplace's instigation a mathematical *tour de force* in which he translated Huygens' construction into algebra, Malus pressed ahead with a careful experimental investigation which showed convincingly that the construction is extremely accurate. Here Malus deployed both the engineering and mathematical training that he had acquired at the École Polytechnique. His work required an acute combination of analysis with cleverly designed and deployed apparatus, yielding in the end what was, at the time, the most accurate optical measurement that had ever been made. Neither he nor Laplace, however, concluded that Huygens' pulse theory of light, which they carefully and thoroughly distinguished from the Newtonian alternative, must therefore be accepted. Instead, both argued, in different but equally peculiar ways, that the resultant formulae are in fact compatible (and perhaps even uniquely compatible, if one takes Laplace at his word) with the mathematics of particles and forces. This debatable claim was quite persuasive among Laplace's associates and students, as well as in certain quarters in England, and for more than a decade and a half optics remained closely bound to the particle theory, as we will see.[3]

Indeed, the persuasive claims of Newtonian optics were furthered in no small measure by Malus' own discovery of an entirely new optical process—the first such discovery since the seventeenth century. Huygens had already noted that light emerging from doubly-refracting crystals seems to have some sort of asymmetry associated with it, since on entry into a second crystal it is not equably divided in two again. In 1809 Malus found, through a series of acute experiments due initially to a serendipitous observation, that this property (which in 1811 he named polarization) did not require a crystal, but that reflection at a particular angle from any transparent body can also produce it. This discovery stimulated a great deal of experimental and theoretical work during the next decade, undertaken especially by Arago and Biot in France, as well as by David Brewster and, somewhat later, John Herschel in Great Britain. Indeed, the most heavily pursued area of quantitative experimental research in optics during the 1810s orbited about the many instrumental novelties and consequent research opportunities opened by Malus' discovery, particularly when the new light form was passed through thin crystal slices, producing beautiful and complicated colours. This work was not based on the Newtonian theory *per se*, nor was it in any easy sense simply instrumental, even though strongly connected to the new device—his polarimeter—that Malus had designed to produce and to measure the new optical property. It was nevertheless hypothetical.

In this scheme the ray of light (not the Newtonian optical particle) provided the fundamental theoretical tool. Practitioners of ray optics, for whom we will shortly introduce a different name appropriated from Thomas Young (the English polymath who, we shall see below, invented a scheme for waves similar to the first one deployed by Fresnel) considered the ray to exist as an individual object that could be counted, and that rays collected together in groups, or bundles, to form beams of light. In this system the ray itself was the central physical object, and the appropriate mathematics involved ray-counting, or what amounted to a species of ray statistics. The character of the system appears strikingly in Malus's own conception of polarization. The intensity of a beam of light is measured numerically by the number of rays that it contains. Unlike the individual ray, which is too weak, a beam can be seen, and its intensity can be manipulated, if not measured directly, using Malus' polarimeter to sort out the rays of different orientations that comprise it, according to the following way of thinking.

Every ray, Malus insisted, has an inherent asymmetry about its length. Think of it rather like a stick to which a crosspiece is nailed at right angles. Given the direction of the ray, the orientation of the crosspiece in a plane at right angles to the ray determines its asymmetry. As Malus understood the concept, 'polarization', properly speaking, does not apply to the individual rays in a beam but only to the beam as a collection of rays. A beam may be polarized in a certain way, but the individual rays that make it up are not themselves said to be polarized, though each has a certain asymmetry. If the asymmetries of the rays in a given beam point randomly in many directions then the beam is, in Malus' understanding, 'unpolarized'. If, on the other hand, one can group the rays in a beam into a number of sets, each of whose elements shares a common asymmetry, or even if this can be done only for a certain portion of the rays in the beam, then the beam is said to be 'partially polarized'. If all of the rays have the same asymmetry then the beam is just 'polarized'. According to Malus' way of thinking, his polarimeter picked out those sets of rays within the beam that had specific asymmetries.[4]

We shall hereafter refer to Malus, and to those who thought like him, as 'selectionists', since they conceived of polarization as a process in which the rays in a beam are selected and have their asymmetries altered in direction. Selectionism was, on the one hand, not at all coincident with the Newtonian theory, since selectionists could and did draw the distinction in controversy with their wave opponents, but, on the other hand, it was nevertheless thoroughly hypothetical, which is easy to see, because on wave principles it is in fact unsustainable—in wave optics light beams cannot be thought of as collections of discrete rays. More to the point, selectionist principles could be, and indeed certainly were, used to develop mathematical laws that had direct application in complex experiments. These laws were neither mere summaries of experimental results, though they were certainly tied directly to particular kinds of instruments, nor were they simply pulled out of the air. On the

contrary, they were deduced directly from the fundamental principles of selectionism, and they are incompatible with laws for the same kinds of phenomena that are implied by the principles of wave optics. However, the two devices (the crystal and the polarimeter) that could be used to examine polarization at the time depended critically upon the eye to judge the presence, absence, or even intensity of light, and in these kinds of experiment the unaided eye has limitations. The limitations were sufficient to preclude any experiment until well into the 1840s that could tell the difference between the selectionist and wave formulae for the most widely influential phenomenon; namely, the partial reflection and refraction of light at the surfaces of transparent media. Indeed, the relationships that Fresnel obtained for calculating the quantities of polarized light reflected and refracted remained without experimental support for decades, thereby generating a pointed controversy concerning them, so long as only the eye could be used to compare optical intensitites.[5] Nor was this the only area in which the difference could not be told.

Malus did not live long enough to develop his own system completely, though its outlines were quite apparent to many people at the time. Arago, for one, saw clearly that Malus' work depended on the division and grouping of light-rays into related sets, and this aided him in explaining a phenomenon involving Newton's rings that he himself discovered shortly before Malus' death. Arago was aged just 23 at the time of his election to the astronomy section of the *Institut de France* in 1809, and his work on Newton's rings two years later, at the age of 25, represented the only research for which he could claim sole responsibility. He had reason to be jealous and intensely proud of his results. From the outset, Arago fully adopted Malus' terminology and understanding of polarization, and with these ideas in mind he decided to make his mark by examining the polarization of Newton's rings. These coloured bands occur when light passes through the narrow gap between, for example, two lenses pressed hard together, and had been extensively investigated by Newton. Working with lenses at the observatory, Arago thought to examine the polarizations associated with the rings. When he did so he rapidly discovered an apparent exception to the rules that Malus had offered for the polarization of reflected light—an exception that Malus himself found to be quite troubling when Arago told him about it. However, as Arago pursued his discovery he did not abandon Malus' understanding, but instead supplemented it by drawing a new distinction between the formation of rings and the generation of their polarization. This, as we shall see in a moment, captured him in a very important way. However, on 11 August he described another, eventually highly influential, discovery he had made that was later termed 'chromatic polarization', involving the generation of coloured patterns by the passage of polarized light through crystal sections.[6]

What happened next proved to be critically important. Between 11 August and the following spring Arago continued to pursue his exciting new discoveries, though we do not know precisely what he was doing during this period. Burdened with heavy teaching and administrative duties he did not, he wrote later, have the time to gather his new work together for a public reading before a disaster occurred. All of a sudden, seemingly out of nowhere, Biot intruded on Arago's field of research and read a note on chromatic polarization that at once thoroughly stripped Arago of his leadership in the new field. Arago demanded that notes he had earlier deposited be examined to show that he had already done what Biot claimed. The Institute appointed Burckhardt and Bouvard to look into this contentious issue, and in April the investigators announced that 'the declaration made by M. Arago [is] most exactly true'. But these notes were actually published only years later, and Arago did not read anything else to the Institute until the following December, eight months after the debacle with Biot over priority. In the meantime the active Biot himself read an extraordinarily long memoir on chromatic polarization to the Institute, followed six months later by an even lengthier discussion that had a major impact on most of his French, and eventually his British, contemporaries.

Arago had lost control of the field he himself had founded. Yet the notes that he had deposited, and which he forced the Institute to examine, scarcely mention the subject of chromatic polarization, which is what concerned Biot. Moreover, Arago's two subsequent memoirs in the general area, as well as his contemporary unpublished notes, are very different in character from Biot's memoirs. Unlike Biot's work, this material is entirely qualitative and yet nearly devoid of any concern with the principles of Newtonian optics. It is almost entirely involved with the overall features of what happens to rays, rather than with why it happens to them or with representing mathematically, in the fashion of Malus, precisely how it happens. But what the notes do contain, though only in a highly undeveloped form, is a general theory that tries to unite the polarization effects of double refraction and thin crystals with those of reflection.

The theory is rather vague and completely non-quantitative, but it does try to unify very different phenomena. Unlike Biot's pre-emptive work in the area, Arago's was very broad in scope. He never attempted to generate formulas from it, and his work does not contain numerical, much less tabular, data of any kind. Biot, by contrast, produced formulas very early on in his work on chromatic effects (though he had nothing like Arago's unifying theory), and his lengthy papers are filled with extensive tables. Of the two, Arago was working in the more traditional, qualitative manner; he was seeking broad principles to encompass several classes of phenomena. Tabular data and formulas do not fit well that kind of endeavour. Biot turned instead to very sharply limited assumptions and pointedly sought to generate formulas from them for specific cases, while attempting to marry his mathematics to quantitative experiment. He made no effort in his early work to link these results to wider classes of phenomena in any firm way.

Biot's first work in this area therefore follows the new pattern of the late 1700s, a pattern that was firmly established in optics by Malus. This pattern was rapidly becoming a standard one. It insisted upon the careful tabular presentation of numerical data and the generation of formulas that are capable of encompassing the material at hand, with little immediate concern to reach out to other, even closely related, phenomena. The differences between Biot's and Arago's work therefore hinged upon changing canons of experimental reporting and investigation, canons that had first appeared in optics in sharpest relief in Malus' work.

During the next several years Biot not only gained fuller control over the subject that Arago had created, but also published long and intricate memoirs that linked it to Newtonian optics. Biot's rapid progress in chromatic polarization culminated in an immensely detailed book, a text that symbolized in concrete form the astonishing success of his endeavour.[7] To many people Biot became that theory's primary exponent. Arago could not have been overjoyed. Even before these events Arago had expressed some doubts about the Newtonian system, or at least about several aspects of it. Little wonder that he came to dislike it violently. Arago was accordingly well prepared to react when in September 1815 he received a long letter from Augustin Fresnel, the nephew of Léonor Merimée (once teacher of design at the École Polytechnique, and by then permanent secretary of the École des Beaux Arts). Fresnel approached Arago with just the kind of quantitative work that Biot could produce and that he, Arago, could not.

Fresnel, himself a graduate of the École Polytechnique a decade before, had briefly visited Arago in Paris in the previous July while on his way to internal exile at his mother's home for having greeted Napoleon's return from exile by joining the Duc d'Angoulême's resistance. Fresnel was already pondering optics by then, and asked about diffraction. Arago gave him a list of English authors, including Thomas Young, which Fresnel could not read, though his brother Leonor could. Fresnel's fall letter to Arago advanced an optical theory similar to the one that had already been discussed by Young, and it contained precise experiments, numerical detail, and beautiful formulas.[8] Most importantly, it seemed to Arago to show something that neither Biot (nor Arago himself) could have predicted: namely, that the coloured fringes produced by light that passes the edge of a narrow object the diffraction fringes move away from the diffractor along hyperbolic paths. Though Fresnel himself did not emphasize this discovery in his first letter to Arago, Arago seized on it and at once pressed Fresnel to improve his observations, to make the discovery indubitable. In passing, he remarked that Fresnel's theory was essentially the same as the Englishman Thomas Young's, though Arago did not at the time also realize that Young had already pointed out the hyperbolic law (which had in any case been remarked for a different configuration than Fresnel's as an empirical generality by Newton in his *Opticks*). What gripped Arago was not so much the excitement of a new discovery as the opportunity to make use of it to redress the recent wrongs he had suffered at the hands of Biot, whereas Fresnel was deeply perturbed by Young's priority, and this in the end stimulated him to even greater exertions. Their two worries nicely intersected for a time.

15.3 Thomas Young and Interference

In 1799 Thomas Young in England had begun a series of publications that substantially extended the quantitative power of medium theories of light. Young, like Leonhard Euler before him, associated colour with wave frequency, but he went far beyond Euler in his use of the assumption (not least because Euler's "frequencies" referred to arithmetically-separated pulses of light).[9] Trained as a medical doctor, Young had during his studies become deeply interested in acoustics, and especially in phenomena of superposition. This eventually led him to the *principle of interference*, according to which continuous waves of the same frequency and from the same source will, when brought together, produce regular spatial patterns of varying intensity. That principle had not been well understood for waves of any kind until Young began his investigations, and it can indeed be said that the study of wave interference in general began with Young himself, though he did not pursue it extensively outside optics. Many difficult problems had to be solved by Young, including the conditions of coherence that make detectable spatial interference possible at all. Furthermore, the principle of superposition, according to which waves combine linearly and which is a necessary presupposition for the principle of interference, was itself quite problematic at the time and also had to be developed and argued for by Young.

Young applied his principle of interference to the diffraction of light by a narrow body, as well as to the case of light passing through two slits, though in the latter case it seems that he did not carry out careful measurements. In all cases he explained the fringe patterns that he observed by calculating the path difference between a pair of rays that originated from a common source. He did not calculate with waves themselves, but rather assigned periodicity to the optical ray, which accordingly retained a signal place in Young's optics. Although Young was certainly quite familiar with Huygens' work, he did not utilize the latter's reduction of rays to purely mathematical artifacts, for that was bound to Huygens' principle, which Young found difficult to accept.

In any event, Young's optics did not generate extensive immediate reaction. Indeed, his principle of interference was sufficiently difficult to assimilate that no other applications to new phenomena were forthcoming. The most famous, or (in retrospect) infamous, reaction, was that of Henry Brougham. Brougham vehemently objected to Young's wave system as an alternative to the Newtonian scheme of optical particles, and he also objected to the principle of interference itself even as a mathematical law applied to rays. Most contemporary optical scientists were more interested in the physics than in the mathematics of light, and in this area Young's ether posed as many qualitative problems as did the alternative system of light-particles. Although Young's work did not stimulate extensive discussion in France, it was nevertheless known there.

15.4 Fresnel, Interference, Diffraction, and Arago

Arago soon brought Fresnel to Paris and participated in new experiments with him, particularly in ones that seemed to hold out the possibility of casting doubt on some

aspect of the Newtonian theory, and so ultimately on the worth of Biot's work.[10] This was the origin of a famous mirror experiment, in which interference occurs between rays that do not pass near material edges and so cannot presumably be affected by the forces that might otherwise be used to explain the formation of fringes. Fresnel's own views underwent considerable development during the next three years, and Arago evidently made certain that he had control over when and where the work was reported. Then, on 17 March 1817, the Academy publicly announced that it had decided to offer a prize on diffraction. By this time Fresnel had extensively changed his original theory, having evolved it from one that was based on comparing rays two at a time to one that was based on wave fronts, Huygens' principle and elaborate integral methods. These developments were stimulated by a succession of increasingly exact experiments in which Fresnel modified his early theory in the face of countervailing observations. His final results reached an extraordinarily high degree of accuracy in placing the loci of diffraction fringes. To do that required Fresnel to develop a series of observational techniques that were designed to provide just the right sort of data for him to deploy numerical methods for approximating his theoretical formulas.

Figure 15.1 is adapted from a diagram drawn by Fresnel himself. In it, C represents the source of a spherically symmetric front AMm' that is intercepted by a screen AG. Adapting Huygens' principle, Fresnel conceived that each point on the front itself emits a spherical wave, albeit with an amplitude that decreases with inclination to the line joining that point to the source C. Introducing z as the distance along AM from the edge A of the diffractor, Fresnel could then represent the amplitude ψ of a disturbance with wavelength λ sent to an arbitrary screen point P in the following way:

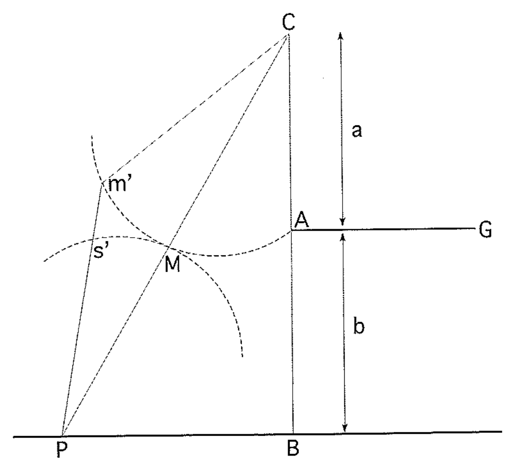

Fig. 15.1. Fresnel's structure for calculating diffraction.

$$\Psi = \sin\left[2\pi\left(t - \frac{CM + m's'}{\lambda}\right)\right]$$

$$m's' \approx \frac{z^2(a+b)}{2ab}$$

whence

$$\Psi = \sin\left[2\pi\left(t - \frac{CM}{\lambda}\right) - \pi z^2 \frac{a+b}{ab\lambda}\right]$$

which decomposes into

$$\Psi = \cos\left[\pi z^2 \frac{a+b}{ab\lambda}\right]\sin\left[2\pi\left(t - \frac{CM}{\lambda}\right)\right]$$
$$+ \sin\left[\pi z^2 \frac{a+b}{ab\lambda}\right]\sin\left[2\pi\left(t - \frac{CM}{\lambda}\right) - \frac{\pi}{2}\right]$$

Fresnel could conclude from this that the square of the resultant from all of the secondaries on the front, pairing up all terms with cosine amplitudes and all terms with sine amplitudes, can be computed from the following sum:

$$\left(\int \cos\left[\pi z^2 \frac{a+b}{ab\lambda}\right] dz\right)^2 + \left(\int \sin\left[\pi z^2 \frac{a+b}{ab\lambda}\right] dz\right)^2$$

or, using a change of variables, to find

$$\int \cos\left(\frac{1}{2}\pi z^2\right) dz \text{ and } \int \sin\left(\frac{1}{2}\pi z^2\right) dz$$

There were two major difficulties with this result with which Fresnel grappled. One was how to establish an appropriately general coordinate system for calculation. This arises in the simplest case when the diffracting object, or aperture, has two edges, for then two limits are involved, and this in effect requires computing values for all points of a surface. The other, which arises in all cases, including that of the semi-infinite plane, where only one boundary occurs (and where, accordingly, the surface just mentioned reduces to a line), is simply how best to calculate useful values for these integrals. Fresnel sought the quickest route to application, and that was by numerical integration (instead, for example, through a series expansion, though Cauchy later developed divergent ones for these integrals). In an astounding computational *tour de force*, Fresnel tabulated the integrals in steps of 0.1 from 0 to 5. His computational errors in doing so amounted to a mean of only 0.0003, and the differences between his values and more accurate ones computed using the series later produced by Augustin Cauchy still amounted to only 0.0006.

The appearance in physical equations of solutions that could only be evaluated by series expansions or by numerical integrations was by this time not altogether unusual. They had emerged quite directly in astronomical problems, and they were soon also to appear in problems involving elasticity and heat flow. Nevertheless, what was unusual was the presence of such things as a *fundamental* expression of the underlying physics. For Fresnel's integrals, unlike, say, Legendre polynomials in astronomy, or, later, Fourier series in thermal processes, were not produced as solutions to given differential equations. On the contrary, they were asserted by Fresnel without his having had in hand the partial-differential equation of which they were meant to be the solutions, much less the methods and techniques necessary to solve such a thing under appropriate sets of boundary conditions. Fresnel, one might say, had discovered the solution to what would later be termed the 'reduced wave equation', or after 1860 as the 'Helmholtz equation', without having any idea at all what that equation was. This character of Fresnel's wave optics contrasted strikingly with that of an optics whose physical imagery at least was grounded on particles and forces, for there the fundamental differential equations were well-known and were, in addition, ordinary.

Of course, there was no way at all to treat what amounted to an n-body problem in Newtonian optics. Consequently, one has the seeming paradox that the only way mathematically to deal with diffraction required a method whose underlying physics and associated fundamental mathematics remained altogether unknown, whereas the alternative physical scheme, in which the physics was very well understood indeed, yielded no effective mathematics at all. Yet this very difference was precisely what aided the assimilation of Fresnel's mathematical methods, and eventually of the physical conceptions that they brought along with them. For no-one on the prize commission, which included Laplace and Siméon Denis Poisson, who were arch-proponents of particle physics, objected that Fresnel's integrals should not be exploited as, in effect, phenomenological expressions of the empirical facts of the matter. During the remainder of the century, Cauchy, George Green, George Stokes, Hermann Helmholtz, Lord Rayleigh, Gustav Kirchhoff, and Arnold Sommerfeld, to mention only a few, grappled with the creation of physics for both sound and light based on the partial-differential wave equation in three-space and its useful solutions. It may with only slight exaggeration be said that all of them were attempting to discover just what these original solutions of Fresnel's were solutions to, and how they could in retrospect be justified or amended.

Laplace, Poisson, and even Biot were not overly disturbed by Fresnel's success in developing formulas for diffraction, not least because they did not think that diffraction was a major topic in optics. Certainly it was an important topic, but in their eyes its application was limited to what happens near edges or when light rays are made to interact with one another in certain kinds of situations. They certainly did not believe that all of optics had to be reconstructed on a new foundation, for they continued to think of the optical ray as a fundamental, irreducible element. They rejected waves. Indeed, acceptance of Fresnel's diffraction formulae without a simultaneous acceptance of Fresnel's wave theory remained quite common until the late 1830s. This does not, however, mean that the Newtonian theory *per se* retained its power. It means rather that a distinction was drawn between that theory and the assumption that rays are individual, countable objects, which remained tacitly unquestioned. Fresnel's work certainly raised many doubts about light-particles, but for quite some time it did not raise corresponding doubts about the physical identity of the light-ray.

15.5 A New Mathematical Optics and the Wave–Particle Debates

Diffraction had not been the exciting topic in optical research throughout most of the 1810s. Polarization, and chromatic polarization in particular, had captured the centre of attention and had made Biot's reputation. This was the subject that Biot had ripped out of Arago's hands in 1811. In mid-July 1816 Arago apparently suggested to Fresnel that he should examine the fringes produced by the interference of the two polarized beams that emerge from a crystal. Fresnel's first discussion of chromatic polarization, which was based on the principle of interference, was not an improvement over Biot's, primarily because at that time Fresnel did not know how to compute the resultant of more than two interfering rays. And though Fresnel submitted this work, which was handed to Arago and Ampère to report on, no report appeared for five years.

Arago's 'report', when it finally did appear, was a polemic directed at Biot. He had delayed five years in writing it probably because it took him that long to understand fully how Fresnel's work might be used to defeat Biot. Biot was given the papers on which Arago was ostensibly reporting, and he saw at once what Arago had done. Remarking that Fresnel himself 'had not proposed as the basic purpose of his work to show that what he calls my theory of mobile polarization is, in many points, insufficient and inexact', Biot concluded by complaining that the report 'deviates from the rules generally established in scientific societies for assuring the equity of their decisions'. The controversy that ensued was nasty, and at times vicious, though Fresnel was himself somewhat on its outskirts. He had written his original memoirs not to attack Biot directly but rather to present his own theory. Arago turned that around, and Biot knew it. However, in the heat of his reply, Biot challenged aspects of Fresnel's work, thereby opening himself to a powerful attack from Fresnel. This confrontation between Biot and Arago, and Biot's subsequent failure to clarify the nature of his theory in exchanges with Fresnel, marks an epoch in the history of optics. Scarcely two years after Fresnel had won the prize for his diffraction memoir, the Institut, now returned in the aftermath of the monarchy's Restoration to its original name as the Royal Academy, ordered printed, over Biot's explicit and public objections, Fresnel's account of chromatic polarization. The members of the Academy did not accept Arago's report, but apparently only because Arago had not insisted that they do so (and he in any case published it almost immediately in his own journal).

Fresnel's final theory of polarization, more than his account of diffraction, broke fundamentally with selectionist optics. As we saw above, contemporary French understanding of polarization considered it to be an essentially static, spatial process in which the rays in a given group have their asymmetries aligned in certain ways. Time does not enter into this scheme, and, according to it, a beam of observably unpolarized light is always just unpolarized, no matter how small a time interval one might consider. For several years Fresnel had found it extremely difficult to discard this notion that common light must not show any signs of asymmetry at all, though he could not use ray-counting procedures to explain why not. He accordingly tried hard to build a scheme in which polarization consists of a temporally fixed combination of directed oscillations: one along the normal to the front, and the other at right angles to it. In common light the transverse component vanishes altogether; in completely polarized light the longitudinal component disappears. This, like its ray-based counterpart, is an inherently static, spatial image. Fresnel was not successful in building a quantitative structure on this basis. Then, in 1821, he set his static image into motion.

The core of Fresnel's new understanding of polarization referred the phenomenon to the change (or lack of change) *over time* of a directed quantity whose square determines optical intensity. This quantity must always lie in the wave front (and is therefore 'transverse' to the ray in optically isotropic media), and in reflection and refraction it can be decomposed, with the components in and perpendicular to the plane of reflection being affected in different ways. Common light consists of a more or less random rotation and amplitude change over time of this directed oscillation, and not, as Fresnel originally understood it, of a spatially fixed (longitudinal) disturbance. Rays of light are mathematical abstractions: they are merely the directions joining the centre of the wave to the front itself. In a purely analytical sense one can link a ray to the asymmetry in the front at the point of the front which the ray contains. Accordingly, in Fresnel's wave theory one can say that a ray is polarized, if one wishes, because polarization refers to the asymmetry at a point in the front, and to each such point there corresponds only one ray. But, and this is a signal characteristic of the scheme, the rays, being mere lines, cannot be counted. A beam of light is not a collection of discrete rays, which means that it cannot be dissected in Malus's fashion. The vast gulf between the two conceptions made it extraordinarily difficult for people who did not think about polarization in the same way to communicate their experimental results to one another without leaving an inevitable residuum of ambiguity that could, and often did, lead to controversies.

Fresnel soon drew startling implications from his new theory of polarization. He produced novel surfaces for doubly refracting crystals that linked polarization to wave speeds and thence to the directions of rays. These surfaces, which Fresnel obtained only after several false starts and much experimental work, opened a new, and unexplored, realm for mathematical and physical analysis, which was soon avidly pursued in England, Ireland, and France by an emerging cadre of wave analysts. Much of this work concentrated on drawing implications from Fresnel's surfaces and associated mathematics. For example, Humphrey Lloyd in Ireland confirmed the existence of a striking phenomenon known as conical refraction that William Rowan Hamilton had deduced; in England, George Airy examined mathematically the special type of double refraction that occurs when light passes through quartz oblique to its optic axis, again deploying with finesse Fresnel's new mathematics for polarization. The late 1830s and the 1840s were also the period during which vigorous debates occurred between partisans of the old and new optics, debates that often raised issues concerning the respective physical foundations of the two theories.

Most wave partisan attacks against the alternative optics looked directly to the latter's putative foundation in particles and forces. Humphrey Lloyd, for example, remarked that the 'emission [Newtonian] theory' of light is essentially useless because 'it is an aggregate of separate principles' concerning the behaviour of optical particles.[11] According to Lloyd, wave optics is nothing like the unamalgamated aggregate of emission (particle) optics, and this is why it can generate new physics. Among the new physics that Lloyd had in mind was partial reflection and refraction. Here, he explained, Biot's physics for optical particles, which Lloyd developed in some detail, simply could not yield formulae—nothing useful emerged from it at all. Fresnel was much better. Unlike emission theorists, who could not generate formulae from their forces and particles, Fresnel could, Lloyd asserted, from the properties of the medium, or ether, in which optical waves were supposed to subsist. Ironically, Fresnel's mathematics for partial reflection and refraction was the one aspect of his new wave theory that escaped experiment for many decades, primarily because the photometric measurements required could be done only by using the eye to judge degrees of illumination, which left considerable scope for argument over results.[12]

Lloyd's brief for the power of the wave theory, and for the intuitive sagacity of its partisans, could not convince anyone on the other side who had spent much time engaged in research, and in fact it did not do so. Neither did William Whewell's remarks in a similar vein produce any conversions, for the fact was that in the early years of wave optics it remained possible still to think in old ways, based on rays of light, because the new system had yet to flower on paper and in the laboratory.[13] During the 1830s, however, wave partisans produced practical analytical tools, and some laboratory tools as well, that enabled them to establish a dynamic research tradition, within which ether physics took its place as one element. Wave partisans, for reasons that did not have much to do with the comparative abstract superiority of their system, also controlled important journals and built a close network of like-minded people through university and (especially) professional associations. Ray partisans did not achieve anything like this; they did not even try to do so. By the 1840s an entire universe of wave devices was being generated—one in which the instruments themselves were increasingly built around the behaviour of wave front and phase, the principal concepts of the new theory. This universe of devices offered no point of entry to the ray physicist, for whom front and phase had no fundamental significance. By then, which was quite some time after the large majority of work in optics had shifted to wave methods, ray physics might be said to be objectively weak in comparison to wave physics.

15.6 The Physical Structure of Wave Optics through Mid-Century

In ray optics the concepts of particle and force served less to generate new mathematics and experiments than to provide a physical foundation. Similarly, in wave optics the ether served for the most part rather to suggest and to justify than to produce new physics. For example, Fresnel's route to a mathematics for diffraction hardly used the ether at all, except for the signal purpose of convincing him in the first place that light did not have to be thought of as a collection of physically distinct rays. In fact, the minor use that Fresnel did there make of the ether prevented him, for nearly five years, from accepting that Huygens' principle had anything useful to say about diffraction. And when he did develop his final theory the ether served him primarily as a strong physical foil with which to counter the physical absurdities (as he saw it) of the optical particle. What it most certainly did not serve him as was a generator of mathematical theory, which in this area at least emerged despite the physical image that Fresnel held of the ether, not because of it.

The same can be said, only even more strongly, of Fresnel's work on polarization. Here he was stymied by his inability to disentangle different types of wave propagation from one another, and to associate only one of these types with polarization. The old way of thinking about polarization—one that Fresnel had much trouble abandoning—linked nicely to the natural image of an optical disturbance as a pressure wave in the ether. The asymmetries that are the core of polarization phenomena required something very different, and the only possibility, Fresnel saw very early on (at Ampère's suggestion), was a transverse wave. This raised problems (such as how the planets could move through such a comparatively rigid thing), but these issues had little effect upon Fresnel in the early years, because even given transverse waves he still did not see how to produce a unified account for polarization. Only after he conceived the entirely novel idea that the asymmetry in the wave front can vary in direction with time as the front propagates did he discover how to construct an entirely new account. And in this development the ether was again more of a stumbling block (as it had been in diffraction) than it was an aid to creation.

But, one might object, there was certainly at least one area where Fresnel did use the ether in much the fashion that the common image of the period suggests—as, that is, a model for producing new mathematical theory: namely, in his creation of the wave surface for biaxial crystals, those which have two optical axes, whereas Iceland spar has only one. His final account, written in 1824, presents an elaborate structure for the ether—one that can, it seems, be used to generate the very difficult series of surfaces that lead to Fresnel's biaxial generalization of Huygens' construction for Iceland crystal. And it was after all this generalization that led, in the hands of Hamilton and Lloyd, to the discovery of conical refraction, which the philosopher William Whewell and others crowed over as concrete evidence of the wave theory's power.

One would nevertheless be wrong to see creative ether dynamics even here, because Fresnel used the image of an elastic ether for precisely one purpose only: to justify physically the assertion that the wave speeds that correspond to an oscillation in a given direction can depend upon that direction. Upon this assertion, and this alone, he built his theory. It led, after several false steps and much hard experimental work, to Fresnel's creation of the surface of elasticity (which is substantially what we now term the optical indicatrix), from which followed the so-called normal surface for front speeds, and thence (through some inspired guessing rather than through rigorous deduction) the wave surface. The ether appears nowhere in this sequence.

However, in his final, published version Fresnel did devote a few pages to suggesting a foundation for ether dynamics—a foundation that he certainly knew to have been flawed, but one that at least seemed to profound a physical basis for the complicated series of steps which derived from the surface of elasticity, a surface that had its physical seat, after all, in ether itself. What, then, was this basis? The ether, for Fresnel (and for many later optical scientists as well) was a system of point masses that exert central, repulsive forces upon one another. This much had almost certainly been in the back of Fresnel's thoughts for many years. But what did he, or even could he, do with this image? Creatively he did scarcely anything at all with it. But he did provide the elements of a foundation for the wave properties of such a system in a way that provides a useful contrast to what the mathematician Cauchy did in the 1830s.

Fresnel developed his ether dynamics on two levels, the highest was closely tied to the immediate demands of optical theory, while the lower was intended to provide a foundation for the upper. The latter asserts essentially two things. First, that the vector reaction to the displacement of any given point is a linear function of the displacement's components. Second, that the coefficients governing the reaction are constants. From this, Fresnel was able rigorously to generate his 'surface of elasticity'. To prove it, however, he made the questionable assumption that the reaction generated by a given point's displacement can be calculated by holding every other point fixed. He was himself entirely aware that the assumption was not a reasonable one, but only by making it could he reach the surface of elasticity from some facsimile of ether dynamics.

This again shows that for Fresnel the behaviour of the ether did not have to be known in order to formulate his new optical theory, including that part of it which does impinge most directly upon the ether's properties: namely, double refraction. But one must not take this too far. Without the ether as a physical underpinning it would have been extremely difficult for contemporary physicists to use the wave theory without feeling that it lacked a firm physical foundation, whereas the alternative to it, which was based on sets of rays, did at least have an apparently firm physical basis in the image of the optical particle governed by forces. Indeed, one of the criticisms that wave theorists often threw at their predecessors was that the ether was a much sounder physical foundation than was the optical particle—in part because, as Fresnel had begun to show, and as Cauchy showed with much greater rigour, one could actually generate from the mechanical structure of the ether theorems that lead to formulas, and not only in double refraction.

It is unlikely that had Fresnel lived he would have continued working much as he had in constructing the wave theory between 1815 and his death in 1827. On the contrary, both published work of his as well as manuscript evidence indicates that the ether was becoming increasingly important to Fresnel as a generator of theory. One phenomenon in particular seemed to him to require an intimate knowledge of the ether's behaviour: the dispersion of light. In 1823 Fresnel advanced a qualitative explanation of dispersion, based on ether dynamics, that became immensely

influential in the 1830s, particularly for Cauchy. Dispersion, according to Fresnel, depended on the spacing and forces between the mutually repelling particles of the ether. The clear implication of Fresnel's remarks was that theory had to address these two factors (spacing and force) in order to deal quantitatively with dispersion. This was precisely what Cauchy had begun to analyse, in a different context, in 1827, the very year that Fresnel died. And Cauchy's mathematics for dispersion set a programme of research that was pursued in France, Britain, and Germany during the 1830s and (in Germany and France) into the 1850s. During the 1830s, in fact, optical theory became for a time nearly synonymous with Cauchy's ether dynamics.[14]

Cauchy's early ether dynamics dates to 1830. By that time he was well prepared to see the possibilities in Fresnel's suggestions, and in particular immediately to correct the great lacuna in Fresnel's dynamics: namely, the assumption that the ether lattice remains essentially rigid even when one of its elements is displaced. Assuming only that the displacement is small in comparison with the distances between the points, Cauchy was able to generate a differential equation in finite differences for the motion of an arbitrary lattice element in function of the differences between its displacement and that of every other element in the lattice. This equation became so common in optics articles and texts during the ensuing decade that it should be called canonical. To produce from these intricate expressions a theory of dispersion whose constants reflected ether properties was no easy task, as the almost impenetrable mound of computations and approximations that Cauchy eventually published in 1836 would seem to show. In essence, Cauchy first imposed symmetry conditions on the lattice and then calculated the differences in the displacements by means of a Fourier series. After a very great deal of tedious work these elephantine calculations birthed what some of his contemporaries considered to be a very small mouse: a series for the refractive index that has since been known eponymously as 'Cauchy's series'. In fact, Fresnel had himself generated precisely the same series, though he had never published it, though in his (unlike Cauchy's) the physical meanings of the associated constants remained undeveloped.[15]

Nevertheless the series did seem to work empirically, though precisely what even that meant was open to question, since it had so many freely disposable constants that at least one English physicist (Samuel Earnshaw) felt that it amounted to an identity. Moreover, there was more than one way to extract a dispersion formula from Cauchy's general structure, and even then it was not clear just what the manipulation of constants meant physically. Still, there can be no doubt but that many of Cauchy's contemporaries, including William Rowan Hamilton, were deeply impressed by his ability to obtain a dispersion formula.

Articles on Cauchy's theory poured forth, particularly in Britain, and a very great deal of thought was devoted to it. But, despite the structure's difficult birth of a dispersion law, it was in other respects almost entirely barren. Insofar as the laboratory was concerned, Cauchy's structure was an alien being. He used Fresnel's wave surface as a sort of intellectual laboratory against which to test its results, but nothing empirically novel ever emerged. It remained an exercise in the highest reaches of abstract model-making, or rather it did so until a moody Irish mathematician named James MacCullagh found its Achilles heel.

Many years earlier, Arago had discovered the phenomenon of optical rotation, in which the plane of polarization of light is rotated on passage through quartz. This discovery had in fact been an important clue for Fresnel in his development of a new understanding of polarization. Cauchy's ether lattice had in principle to embrace all optical phenomena, including this one, the procedure being to effect suitable adjustments of the constants that describe the state of the lattice in equilibrium. The equations of the structure are so complicated and difficult to penetrate with any degree of ease that this gargantuan claim seemed entirely plausible, particularly in view of Cauchy's novel success for dispersion, and his early success for double refraction. But in 1841 MacCullagh unequivocally demonstrated that the very structure of the equations forbids optical rotation: the kind of lattice—or better put, the kind of mathematics for the lattice that Cauchy had created—fails without hope of salvation to capture a very important phenomenon indeed.

As far as MacCullagh was concerned, this ended what he in any case thought to be the building of fanciful structures to encompass things that are already known perfectly well (since, like Earnshaw, MacCullagh did not consider the dispersion law to be much more than a sort of interpolation, something with about the same empirical force as a Taylor series expansion). In Britain, Cauchy's structure did indeed disappear rapidly, but not simply because MacCullagh had dealt it the analytical equivalent of a mortal blow. In fact, the structure could be salvaged; Cauchy himself never had any doubt that it could be, and his several French (and later German) supporters agreed with him. Indeed in 1847, six years after MacCullagh's mathematical tirade, a French abbé by the name of Moigno produced a hefty *Repertoire d'Optique Moderne* which was essentially a long argument for the wonders that Cauchy had produced—one that simply treated optical rotation as a troubling little problem that could be made to go away. The problem could be overcome, but only by creating a new mathematical structure for the lattice that made Cauchy's earlier equations seem the soul of simplicity by comparison.

To solve the problem Cauchy in effect created the mathematical theory of periodic structures using concepts that had recently been introduced by Auguste Bravais in crystallography. In terms of his old equations, what this meant was that the 'constants' had now to be treated as periodic functions of position. Cauchy was able to demonstrate, after a truly acrobatic performance with the vast array of relations that periodicity provided, that there are new low-order terms that can capture optical rotation. This took place in 1850, nearly a decade after MacCullagh's critique. By that time the early, powerful influence of Cauchy's conception of the ether lattice had vanished entirely in Britain and was becoming less common in France, though it continued to crop up in Germany, which was only just then beginning to contribute in substantial numbers to research in mathematical optics.

The physical models that gripped theorists of the optical ether, as well as the problems that they addressed, have their origins in Fresnel's work, as we have seen, but here we find for the first time physicists attempting to grapple quantitatively with the models themselves rather than relying exclusively on the more fundamental principles that the models were designed to encompass. In this respect the period of the 1830s during which Cauchy's structure was most avidly pursued resembles somewhat the late 1890s and early 1900s, during which quantitative models for the microstructure of matter were first built that could yield testable optical formulas. Indeed, there was no earlier attempt by physicists to develop on this scale the implications of a detailed model for things that cannot be observed. Until the advent of the electron, these kinds of investigation remained unusual and even highly controversial.

By the early 1840s at the latest there were few physicists or mathematicians who disputed the wave theory's fundamental principles, and of equal importance, many were by then capable of applying it at the high level of mathematical detail that it required. During this period, investigations based on the wave theory began to evolve into two related but distinct areas. Work designed, like Cauchy's, to pursue the implications of ether mechanics continues. However, the fact is that by the early 1850s very few published papers in optics directly concern mechanical deductions; rather, the vast majority of them involve the working out of the wave theory's mathematical principles and their applications.

Ether mechanics proper has received the lion's share of historical attention, though it really has the mouse's share of contemporary optical production. Relativity has so thoroughly coloured modern views that practically the entire Victorian era has often been treated as a sort of prelude to it. This is incorrect, if only because noone in the nineteenth century knew what was coming in 1905. This is why Cauchy's ether mechanics was something of an aberration rather than a central part of optics: ether mechanics was simply not at the centre of most optical work during the period, though there were times when it did move closer in as it seemed to promise something useful. The preponderance of work concerned such questions as the proper form of the solutions to the wave equation, how to build optically significant equations for various kinds of media, and so on—some of which did touch on the ether, but most of which was substantially independent of it.

There is, however, no doubt but that a considerable amount of powerful effort was devoted to ether mechanics, particularly before 1870 or so, though after the early 1840s a second trend emerged that differed considerably from Cauchy's. It emerged first in the work of James MacCullagh and of George Green, followed in the 1850s and 1860s by that of George Stokes and others. Here lattice equations played no role at all; indeed, they were conspicuously avoided. Instead, the ether's equations were obtained by manipulating a macroscopic potential function, which took on energetic significance after the 1850s and which (in the case at least of Stokes) could be motivated by considerations of the microworld. These two methods—what one might call the molecular and the macroscopic—shared the assumption that matter affects the ether by altering the coefficients that determine its equations of motion (though the theories diverge from one another over whether to alter density or elasticity coefficients).

However, even during the 1840s some physicists thought this assumption was inadequate and believed ether and matter had to be treated as distinct systems that are dynamically connected to one another. This way of thinking disappeared, or at least became a conviction rather than a programme, until the discovery of anomalous dispersion in 1870, on which more below.

15.7 Optical Problems: Huygens' Principle, the Diffraction Integral, and the Nature of Unpolarized Light

Although Fresnel had gone far in developing the principles of wave optics, as well as its mathematical foundation, many thorny issues remained. There were, as we just saw, problems concerning the physics of the ether itself, problems that were dealt with in different ways throughout the century. Other difficulties, which were even more fundamental, stimulated considerable research. Perhaps the central problem left by Fresnel in the foundations of optics concerned Huygens' principle. We saw above how he used the principle to produce the integrals that he then applied with great success to diffraction. However, Fresnel's integrals were asserted by him without his having had in hand the partial-differential equation that they were meant to be solutions of, much less the methods and techniques necessary to solve such a thing under appropriate sets of boundary conditions. Even worse, Fresnel had been forced to conjecture, rather than to deduce, the form of the factor that governs the contribution of a given secondary wavelet to the succeeding front along a specific direction (which was later termed the 'inclination factor'). Moreover, Poisson had challenged the very foundation of Fresnel's integrals—his particular use of Huygens' principle.[16] All these difficulties derive from the fact that Fresnel did not possess a partial differential equation for wave optics, much less a method for obtaining its general solution.

One might say with some justification that the problem Fresnel faced had its origins in the uncertain physical underpinnings of wave optics, but in fact, the major difficulty was essentially mathematical. Although the three-dimensional wave equation had been developed during the eighteenth century, its solutions had not been treated in a form suitable for questions concerning the general behaviour of entire wave fronts, not least because the mathematics for boundary conditions remained undeveloped. Poisson was the first to investigate these sorts of problems. Starting with a closed surface at a given time, and with a given distribution of velocity and condensation, Poisson deduced from the wave equation an expression for the disturbance at any point of space and at any time in terms of an integral over the originally specified surface. In effect, Poisson examined the case of an isolated, spherically symmetric

pulse, and he specified particular conditions that it must satisfy. These conditions, it seemed to him, were not easily satisfied by the demands Fresnel placed on Huygens' principle. Poisson admitted the empirical cogency of Fresnel's diffraction integrals, but he rejected Fresnel's justification for them.

Poisson's work on the wave equation accordingly had nothing to offer wave optics—quite the contrary, it seemed to pose problems for it, not the least of which was Poisson's further claim that Fresnel's inclination factor was itself highly problematic. In 1849 Stokes addressed this latter problem, which, ironically, he attacked by using the very solution that Poisson had himself developed. Stokes in fact attempted to develop a theory for the diffraction of a vector wave and from it to deduce the very same inclination factor that Fresnel had long before conjectured, and that Poisson had criticized.[17] However, Poisson's solution was not suitable for addressing the general questions raised by Fresnel's use of Huygens' principle. For his solution began with an aperiodic front and found subsequent fronts by integration over this initial one. Huygens' principle as used in diffraction theory, however, requires considering an infinitely long, periodic disturbance. Indeed, arguments based on periodicity lay at the heart of Fresnel's analysis, and were precisely what Poisson refused to accept.

The route to a useful, general method of front integration, based on the solution properties of the wave equation, was first developed by Hermann Helmholtz in 1860.[18] To analyse the behaviour of organ pipes, he assumed periodicity and reduced the equation to the form $\nabla^2 w + \left(\frac{2\pi}{\lambda}\right)^2 w = 0$ (wherein λ represents the wavelength). The critical step in Helmholtz's analysis involved his use of Green's theorem, which permitted him to express the disturbance at a point as integrals over a bounding surface; these integrals contained the values of the function and its normal derivatives over the surface. Helmholtz's discussion was limited to spherical, harmonic waves, and he did not extend it to diffraction since he was not considering optical disturbances. As a result, difficult questions concerning the boundary conditions to be applied under these circumstances did not arise.

In 1882 Gustav Robert Kirchhoff applied Helmholtz's use of Green's theorem to wave optics. Kirchhoff first extended Helmholtz's result to the general case of an arbitrary, infinitely long disturbance, developed surface integrals suitable for diffraction theory, and then specialized to the case of a purely harmonic form.[19] Kirchhoff's formulae, however, required assuming inconsistent boundary conditions over the diffracting surface, and this inconsistency generated a great deal of subsequent discussion, culminating in a reconsideration of the problem on the basis of electromagnetic theory by Arnold Sommerfeld in 1896, though issues concerning Kirchhoff's analysis remain to this day. Sommerfeld's theory, which made special use of Riemannian spaces, succeeded well for diffraction by perfectly reflecting screens, but still left problems for black, or perfectly absorbing ones. These several analyses, as well as others of the period, reduced Huygens' principle to, at best, a statement concerning the differential terms that appear in general diffraction integrals. It had been stripped nearly altogether of direct physical significance. In the late 1930s, however, B. B. Baker at the University of London and E. T. Copson at the University of St Andrews published an extensive discussion that attempted to retrieve, insofar as possible, the physical content of the principle.[20]

Huygens' principle and the diffraction integrals were not the only problems that troubled wave scientists as they sought to develop and to expand the dominion of the system. Polarization itself posed particularly hard problems. In Fresnel's new optics, polarized light consisted of a directed quantity, located in the wave front proper, whose magnitude and direction change in a calculable manner over time as the front moves through space. To provide an appropriate mathematics for this entity, Fresnel employed a method of orthogonal decomposition: any disturbance could be specified by providing the parameters that characterize its components along a given pair of mutually orthogonal axes. In general, each component has the form $a \cos(\omega t + \varphi)$, where a specifies the component's amplitude and φ its phase. This worked extraordinarily well for handling all the forms of polarized light known in Fresnel's day, as well as for predicting the character of a form that had not yet been investigated (namely, elliptically polarized light).

This system could not, however, easily deal, in the form introduced by Fresnel, with unpolarized or even partially polarized light, because these kinds of polarization do not have stable values of amplitude and phase. Indeed, in Fresnel's optics unpolarized light is defined as light whose phase at least varies randomly over time, while partially polarized light must in some fashion have restricted variations. In neither case can one directly employ for purposes of analysis the decomposition that Fresnel had introduced for polarized light.

In modern terms, Fresnel constructed wave optics on the basis of an amplitude formulation, in which the amplitude and phase of the wave are used directly for purposes of analysis. Yet amplitude and phase cannot be detected directly, they can only be inferred from experiments that work with what can be observed, namely optical intensities and angles. Since unpolarized and partially polarized light do not have stable amplitudes and phases, any system that seeks to incorporate them into a consistent general scheme must work with intensities and angles, and not with amplitudes and phases, with what may be called an intensity formulation. Developed by George Gabriel Stokes in 1849, the new system characterized polarization in terms of four parameters, each of which could be directly observed. The Stokes parameters did not, however, have much contemporary significance (and were not for that matter pursued extensively by Stokes himself) since at the time there were no pressing physical questions to which they could be applied.[21]

15.8 Optics after 1840

Geometric representations of complex numbers had been developed in 1797 by the Norwegian Caspar Wessel, and in 1806 by a Swiss, Jean-Robert Argand. They had not, however, found a natural use in many physical problems. Indeed, wave optics was the first area of physics in which complex numbers provided specific benefits that could not otherwise be obtained. They appeared for the first time in Fresnel's 1823 analysis of internal reflection. When light strikes within, for example, a glass prism placed in air beyond a specific angle of incidence, then it is totally reflected. Fresnel's equations for the ratio of light reflected to light incident yield complex numbers precisely at and above this incidence. In seeking to interpret this apparent failure of his expressions to retain physical meaning, Fresnel developed the first productive (albeit hardly rigorous) use for such things.

He reasoned in the following manner. First of all, Fresnel remarked, the complex expression that he did obtain under these circumstances has the characteristic that the sum of the squares of its real and imaginary parts is equal to the square of the incident intensity. This suggested to him the following interpretation. The light must be totally reflected, and it undergoes some phase shift as well. Using the general decomposition that Fresnel had originally produced for diffraction, such a phase shift can be understood by separating the resultant wave into two parts that differ in phase from one another by 90°. One of the parts has the same phase as the incident wave, and the amplitudes of the parts must be the cosine and sine of the phase of their resultant. In the case of total internal reflection there are two terms (one imaginary, the other real) whose real squares sum to 1. Suppose, Fresnel reasoned, that each of these terms represents the amplitude of one of the parts of the usual quarter-wave decomposition. Then the light will be completely reflected, and it will also be shifted in phase by an angle whose tangent is the ratio of the real to the imaginary part of the complex expression for the reflected amplitude, if we assume that the real part is the one that has the same phase as the incident wave.

In subsequent years Fresnel's interpretation was taken up and further developed by, among others, Cauchy, who himself made fundamental contributions to complex function theory. Cauchy analysed reflection from metals under the assumption that, in them, the index of refraction itself becomes complex, but that the usual Fresnel expressions for the reflected amplitudes remain the same in form. The Irish mathematician James MacCullagh independently hit on the same idea. The Frenchman Jules Jamin was the first to compile tables of the metallic constants for various metals and wavelengths. This approach reached its fullest mathematical form in the German Friedrich Eisenlohr's treatment of the subject, and his equations were widely used during the latter part of the nineteenth century.[22]

The topic of metallic reflection raised several issues that puzzled many scientists during the last quarter of the nineteenth century. Extensive experimental work by, among others, Georg Quincke had shown that for all known metals the real part of the index of refraction must decrease with frequency, implying that metals should (were it observable) exhibit precisely the opposite dispersion of ordinary transparent bodies.[23] Nor was this the only difficulty, for the same measurements also indicated that the wave speed in metals should be more than five times greater than the speed of light *in vacuo*, which raised a number of questions concerning how the specific characteristics of matter affected optical phenomena. Our account of wave optics can close appropriately with a brief discussion of the major changes that began to occur in the

1870s and 1880s, as scientists in Germany shifted their attention from the behaviour of the optical ether itself to issues concerning the interaction between ether and matter.

In 1870 the Danish scientist Christiansen observed a phenomenon that came to be known as 'anomalous' dispersion. Using the aniline dye fuchsin dissolved in alcohol, he determined that the refractive index of the solution increases from the B to D spectral lines, decreases from D to G, and then increases again after G.[24] The existing interest, in Germany, in metallic reflection fuelled concern with Christiansen's discovery, for the peculiar optical properties of metals could be related to his anomalous dispersion. Indeed, scarcely two years after the discovery the German Wolfgang Sellmeier developed a quantitative mechanical theory for it according to which all forms of dispersion are due to the interaction of ether vibrations with the natural oscillatory frequencies of material molecules. The guiding ideas of his account were that the wave equation of the ether is itself unaffected by the presence of matter particles, but that energy must be abstracted from an ether wave in order to displace the massy particle, which is tied elastically to a fixed location. Employing an analysis based solely on energy considerations, Sellmeier deduced a formula for dispersion that yielded the major features of the phenomenon.[25]

The critically important aspects of Sellmeier's theory were its tacit assumptions that neither the elasticity nor the density of the ether itself should be manipulated, its properties remaining effectively fixed, though Sellmeier was not dogmatic on this point. For him these assumptions were primarily conveniences, for he held that ether and matter together actually constitute a dual lattice of point masses. All refractive phenomena then implicate a mechanical resonance action in which ether waves must move massy particles of matter as well as the substance of the ether itself. What Sellmeier's theory lacked was a mathematical representation of what occurs at resonance between the frequency of an ether wave and that of a material particle, which is precisely where absorption takes place.

In 1875 Helmholtz appropriated Sellmeier's basic ideas and reformulated them mathematically in mechanical equations of motion that were capable of dealing with absorption as well as dispersion. Helmholtz's equations were based on the commonly received idea that optical absorption involves the transformation of light energy into the 'inner, irregular motion of the molecules' of matter: that is, into heat. To effect this transformation a force is necessary, and this is what was missing in Sellmeier's theory. Helmholtz's account formed the basis for a great deal of subsequent optical theory, eventually appearing (in changed form) in early electromagnetic optics as well.[26]

Helmholtz was well aware that the fact of absorption (the exponential decrease in light amplitude with distance) requires the presence in differential equations of a term proportional to a velocity (more generally, to an odd-order time derivative), but he also knew that this term cannot appear in the ether's equation of motion since, *in vacuo*, no absorption occurs. Helmholtz's solution to the problem was at once obvious and unprecedented: he constructed a distinct equation of motion for matter that itself contains a velocity-dependent force. The idea was that the material particles, driven by ether waves, are subject to two other forces that emanate from the surrounding matter particles: an harmonic force of restitution and a frictional force of resistance. As the matter particles absorb energy from the ether waves, the optical energy decreases, and the absorbed energy is converted into thermal motion by the frictional force. Just how the latter transformation occurs was foreign to Helmholtz's theory. Neither of the two material forces were supposed to act on an entire molecule. Rather, each molecule consists of a massive central core, which hardly moves when struck by an ether wave, together with a light, moveable particle; the latter is resisted frictionally in its motion, and the harmonic forces tie it elastically to the massive core.

The mutual actions of ether and matter that cause energy transformations between them must, Helmholtz reasoned, satisfy the principle of action and reaction: whatever force represents the action of matter on ether in the latter's equation of motion must appear with the opposite sign in the former's equation of motion. Helmholtz assumed on mechanical grounds that this mutual action is directly proportional to the difference between the displacements from equilibrium of ether and the light, moveable particle of matter. He further treated both media as effectively continuous and interpenetrating, so that the ethereal and material displacements are continuous functions of time and distance. This gave him partial differential equations. The equation of motion of the ether consists of the usual one for an incompressible, isotropic, elastic continuum, to which a term is added to represent the action of matter upon ether. The material equation of motion contains three forces: the ether term (with reversed sign), the frictional force, and the harmonic action. These *twin equations* constitute the mathematical structure of Helmholtz's theory, for they lead at once to a wave equation that is easily applied to all forms of dispersion and to absorption—one that agreed well with laboratory results when proper choices were made for the various constants, as Helmholtz showed in some detail.

Helmholtz's theory was immensely influential, and not only in Germany. During the next fifteen years, numerous German physicists, including Eduard Ketteler, Eugen Lommel, and Woldemar Voigt, used it in one way or another to construct mechanical theories of phenomena in physical optics. These theories, like Helmholtz's, generally gave little detailed consideration to the actual molecular structure of matter, preferring instead to employ simple and usually *a priori* terms in the material equations.[27] Among the British, Lord Kelvin (1884), in his acclaimed *Baltimore Lectures*, based most of the intricate mechanical models for which he soon became famous directly on the Helmholtz–Sellmeier model. Kelvin's work, however, differed substantially from contemporary German accounts in that his goal was to construct a continuum representation for ether and matter.

Helmholtz's twin equations, and the German and British treatments of them before the 1890s, are purely mechanical. By 1878 the idea underlying the twin equations had already been used in electromagnetic theory. H. A. Lorentz in the Netherlands had tentatively supposed that the ether's properties are in themselves invariant and that optical effects are due to the effect of inner electrical motions of material particles. He reaffirmed this idea when he deduced for the first time an electromagnetic formula for dispersion by actually constructing two linked sets of equations: one for the ethereal polarization, and the other for the motion of a moveable charge in a molecule with a fixed central core. The links between the two equations were, on the one hand, the polarization in the invariant ether effected by the moveable charges, and on the other, the driving force exerted by the ether polarization on the charge.[28]

The correspondences between this and Helmholtz's (1875) mechanical theory are manifest. However, Lorentz viewed the ether as itself a polarizable substance, with the result that he drew no clear distinction between material and ethereal polarization. In any case, Lorentz's 1878 theory apparently had comparatively little influence, except perhaps in Holland, for it involved the sort of detailed microphysical computations and presuppositions which, even as late as the mid-1890s, few German physicists were willing to employ. In order for Helmholtz's twin-equation approach to acquire an electromagnetic significance that most German physicists could easily grasp and approve, the equations had to be reinterpreted in a way that preserved both their formal structure as linked systems and their relative independence of detailed microphysical calculations. This was accomplished by Helmholtz himself. His new interpretation worked with Lorentz's to stimulate a thoroughgoing transformation in the foundations of optical science—one that increasingly sought explanations in the structure of molecules and eventually of atoms themselves.[29]

Bibliography

Baker, B. B., and Copson, E. T. (1939). *The Mathematical Theory of Huygens' Principle.* Oxford: Clarendon Press.

Beer, August (1853). *Einleitung in die Höhere Optik.* Braunschweig: Friedrich Vieweg.

Biot, Jean Baptiste (1814). Recherches expérimentales et mathématiques sur les mouvements des molécules de la lumière autour de leur centre de gravité. Paris.

Buchwald, Jed Z. (1980a). 'Experimental investigations of double refraction from Huygens to Malus', *Archive for History of Exact Sciences* 21: 245–278.

——(1980b), 'Optics and the theory of the punctiform ether', *Archive for History of Exact Sciences* 21: 245–278.

——(1989). *The Rise of the Wave Theory of Light: Optical Theory and Experiment in the Early Nineteenth Century.* Chicago: University of Chicago Press.

——(2012). 'Cauchy's theory of dispersion anticipated by Fresnel,' in *A Master of Science History, Archimedes,* 30: 399–416.

Cantor, Geoffrey (1983). *Optics after Newton: Theories of Light in Britain and Ireland, 1704–1840.* Manchester: Manchester University Press.

Cauchy, Augustin Louis (1839). 'Sur la quantité de lumière réfléchie, sous diverses incidences, par les surfaces des corps opaques, et spécialement des métaux,' *Comptes Rendus* 8: 553, 658, 961.

Chappert, André (1977). *Etienne Louis Malus (1775–1812) et la théorie corpusculaire de la lumière.* Paris, Vrin.

——(2004). *L'Edification au XIXe Siècle d'une Science du Phénomène Lumineux.* Paris, Vrin.

Chen, Xiang (2000). *Instrumental Traditions and Theories of Light: The Uses of Instruments in the Optical Revolution.* Dordrecht: Kluwer.

Christiansen, C., (1870). 'Ueber die Brechungverhältnisse einer weingeistigen Lösung der Fuchsin; brieflicher Mittheilung.,' *Annalen der Physik und Chemie* 141: 479–480.

Dalmedico, Amy Dahan (1992). *Mathématisations. Augustin-Louis Cauchy et l'École Française.* Paris: A. Blanchard.

Dijksterhuis, Fokko Jan (2004). *Lenses and Waves. Christiaan Huygens and the Mathematical Science of Optics in the Seventeenth Century.* Dordrecht: Kluwer.

Eisenlohr, Friedrich (1858). 'Ableitung der Formeln für die Intensität des an der Oberflächen zweier isotropen Mittel gespiegelten, gebrochenen und gebeugten Licthes,' *Annalen der Physik und Chemie* 104: 346ff.

Helmholtz, Hermann von (1859). 'Theorie der luftschwingungen in Röhren mit offenen Enden,' *Journal für reine und angewandte Mathematik* 57: 1–72.

——(1875). 'Zur Theorie der Anomalen Dispersion,' *Annalen der Physik und Chemie* 154: 582–596.

——(1893). 'Electromagnetische Theorie der Farbenzerstreuung,' *Annalen der Physik und Chemie* 48: 389–405.

Jamin, Jules (1852), 'Memoir on metallic reflection,' in *Scientific Memoirs,* ed. Richard Taylor, Vol. 5. London: Richard and John Taylor, pp. 66–101.

Ketteler, E. (1885). *Theoretische Optik gegründet auf das Bessel-Sellmeiersche Princip.* Braunschweig: Friedrich Vieweg.

Kipnis, Nahum (1991). *History of the Principle of Interference of Light.* Basel: Birkhäuser Verlag.

Kirchhoff, Gustav Robert (1891). 'Zur Theorie der Lichtstrahlen [1882]', in *Gesammelte Abhandlungen (Nachtrag),* ed. L. Boltzmann. Leipzig: Barth, pp. 22–53.

Kline, Morris (1972). *Mathematical Thought from Ancient to Modern Times.* New York, Oxford University Press.

Lloyd, Humphrey (1834). 'Report on the progress and present state of physical optics,' *British Association for the Advancement of Science Reports*: 295–413.

Lorentz, H. A. (1935–39a). '[1875] Sur la théorie de la réflexion et de la réfraction de la lumière', in *H. A. Lorentz: Collected Papers,* ed. P. Zeeman and A. D. Fokker, vol. 1. The Hague: Nijhoff, pp. 193–383.

——(1935–39b). '[1878] Concerning the relation between the velocity of propagation of light and the density and composition of media', in *H. A. Lorentz: Collected Papers,* ed. P. Zeeman and A. D. Fokker, vol. 2. The Hague: Nijhoff, pp. 1–119.

——(1935–39c). '[1892] La théorie électromagnétique de Maxwell et son application aux corps mouvants', in *H. A. Lorentz: Collected Papers,* ed. P. Zeeman and A. D. Fokker, vol. 2. The Hague: Nijhoff, pp. 165–343.

MacCullagh, James (1880 [1843]). 'On the laws of metallic reflexion, and on the mode of making experiments upon elliptic polarization', in *The Collected Works of James MacCullagh,* edited by J. H. Jellett and S. Haughton. Dublin: Hodges, Figgis, pp. 230–248.

Poisson, Siméon Denis (1866–70). '[1823] Extrait d'un mémoire sur la propagation du mouvement dans les fluides élastiques,' in *Augustin Jean Fresnel: Oeuvres Complètes,* ed. E. Verdet and H. de Senarmont, vol. 2. Paris: Imprimerie Impériale, pp. 192–205.

Quincke, Georg (1866–72). 'Optische Experimental-Untersuchungen', *Annalen der Physik und Chemie,* many issues.

Sellmeier, Wolfgang (1872). 'Ueber die durch Aetherschwingungen erregten Körpertheilchen und deren Ruckwirkung auf die ersten, besonders zur Erklärung der Dispersion und ihrer Anomalien', *Annalen der Physik und Chemie*: 399–421, 520–49 (145); 386–408, 525–44 (147).

Stokes, George Gabriel (1849). 'On the dynamical theory of diffraction,' *Transactions of the Cambridge Philosophical Society* 9: 1–62.

——(1901 [1852]). 'On the composition and resolution of streams of polarized light from different sources', in *Mathematical and Physical Papers,* vol. 3. Cambridge: Cambridge Unviersity Press, pp. 233–258.

VERDET, E., and SENARMONT, H. DE (eds.) (1866–70). *Augustin Jean Fresnel: Oeuvres Complètes.* 3 vols. Paris: Imprimerie Impériale.
WHEWELL, WILLIAM (1837). *History of the Inductive Sciences from the Earliest to the Present Time.* London: John W. Parker.
YOUNG, THOMAS (1855). *Miscellaneous Works.* London: John Murray.
SHAPIRO, ALAN E. (1973). 'Kinematic optics: a study of the wave theory of light in the seventeenth century', *Archive for History of Exact Sciences* 11: 134–266.

NOTES

1. On seventeenth-century wave optics, see Shapiro (1973) and on Huygens, Dijksterhuis (2004).
2. On which, see Cantor (1983).
3. On Malus' discoveries see Buchwald (1980a) and Chappert (1977).
4. On Malus' theory see Buchwald (1989), chap. 2.
5. On which, see Chen (2000).
6. On Arago and the ensuing controversy with Biot, see Buchwald (1989), chaps. 3–4.
7. Biot (1814).
8. This letter, as well as Fresnel's works discussed below, are all reprinted in Verdet and Senarmont (1866–70).
9. Young's papers are collected in Young (1855). On his optics, see Kipnis (1991).
10. Full references to the original literature referred to in this and the succeeding two sections can be found in Buchwald (1989).
11. Lloyd (1834).
12. On which, see Chen (2000).
13. Whewell (1837).
14. On Cauchy's work in optics, see Buchwald (1980b) and (2012); Dalmedico (1992).
15. For a simplified, canonical deduction of the dispersion equations, see Beer (1853), pp. 187–215. Beer, in fact, generated the 'Cauchy' dispersion formula much as Fresnel had in his unpublished manuscript.
16. Poisson (1866–70).
17. Stokes (1849).
18. Kline (1972), pp. 693–694 on Helmholtz (1859).
19. Kirchhoff (1891).
20. Baker and Copson (1939).
21. Stokes (1901) [1852].
22. Cauchy (1839), Eisenlohr (1858), Jamin (1852), MacCullagh (1880).
23. Quincke (1866–72).
24. Christiansen (1870).
25. Sellmeier (1872).
26. Helmholtz (1875).
27. For an extensive contemporary account, see Ketteler (1885).
28. Lorentz (1935–39a), Lorentz (1935–39b).
29. Helmholtz (1893), Lorentz (1935–39c). For a synoptic overview of nineteenth-century optics, see Chappert (2004).

CHAPTER 16

THERMAL PHYSICS AND THERMODYNAMICS

HASOK CHANG

16.1 INTRODUCTION

Heat is a subject that has commanded people's attention through the ages, for practical as well as scientific reasons. It is still a major subject in introductory physics textbooks and courses, though the science of heat is now presumed reduced to classical or quantum-mechanical principles through statistical reasoning. This chapter covers the development of the physics of heat while it existed as a truly independent subject, which is to say, up to the mid-nineteenth century. Most attention will be paid to the important yet relatively neglected parts of the history, while well-known areas will be covered briefly with references to existing secondary literature.

The study of heat began to flourish in the late eighteenth century, particularly in the chemical communities of Scotland and France. Intense theoretical and experimental activity continued in this field in the first half of the nineteenth century, mostly in the tradition of the material theories based on the basic assumption that heat was (or at least could be conceptualized as) an all-pervasive, weightless and elastic fluid, most commonly called 'caloric'. Great advances were made in caloric-based theoretical treatments of thermal phenomena, which became more quantitative and systematic. Experimental knowledge developed continually both in extent and precision, often quite independently of theory. Also significant among the nineteenth-century developments was the relocation of the study of heat from chemistry to physics, partly prompted by the increasing interest in heat engines. In this chapter we survey some of the significant themes in the development of thermal physics up to the establishment of classical thermodynamics. Much of the early achievement in this field was lost when the assumptions of the existence and conservation of caloric were rejected