Like a Bridge over Colored Water: A Mathematical Review of *The Rainbow Bridge: Rainbows in Art, Myth, and Science*

Reviewed by John A. Adam

The Rainbow Bridge: Rainbows in Art, Myth, and Science
Raymond L. Lee Jr. and Alistair B. Fraser
Pennsylvania State University Press , 2001
393 pages, $65.00
ISBN 0-271-01977-8

This is a magnificent and scholarly book, exquisitely produced, and definitely not destined only for the coffee table. It is multifaceted in character, addressing rainbow-relevant aspects of mythology, religion, the history of art, art criticism, the history of optics, the theory of color, the philosophy of science, and advertising! The quality of the reproductions and photographs is superb. The authors are experts in meteorological optics, but their book draws on many other subdisciplines. It is a challenge, therefore, to write a review about a book that contains no equations or explicit mathematical themes for what is primarily a mathematical audience. However, while the mathematical description of the rainbow may be hidden in this book, it is nonetheless present. Clearly, such a review runs the risk of giving a distorted picture of what the book is about, both by "unfolding" the hidden mathematics and suppressing, to some extent, other important and explicit themes: the connections with mythology, art, and science. To a degree this is inevitable, if such a weighted metric is to be used. Lee and Fraser are intent on exploring bridges from the rainbow to all the places listed above, and in the opinion of this reviewer they have succeeded admirably. The serious reader will glean much of value, and mathematicians in particular may benefit from the tantalizing hints of mathematical structure hidden in the photographs and graphics. Following an account of several themes present in the book that I found particularly appealing, some of the mathematical structure underlying and supporting the bridge between the rainbow and its optical description will be emphasized in this review.

I have included several direct quotations to illustrate both the writing style and the features of the book that I found most intriguing. Frankly, I found some of the early chapters harder to appreciate than later ones. This is a reflection, no doubt, of my own educational deficiencies in the liberal arts; but upon completing the book, I was moved to suggest to the chair of the art department that it might be interesting to offer a team-taught graduate seminar in art, using this book as a text. He is now avidly reading my copy of the book.

The ten chapters combined trace the rainbow bridge to "the gods" as a sign and symbol ("emblem and enigma"); to the "growing tension between scientific and artistic images of the rainbow"; through the inconsistencies of the Aristotelian description and beyond, to those of Descartes and Newton, and the latter's theory of color; to claims of a new unity between the scientific and artistic
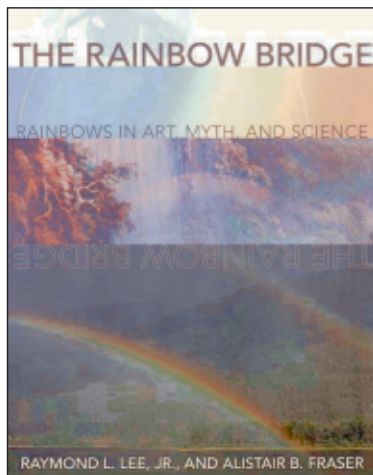
John A. Adam is professor of mathematics at Old Dominion University. His email address is jadam@odu.edu.

enterprises; the evolution of scientific models of the rainbow to relatively recent times; and the exploitation and commercialization of the rainbow. All these bridges, the authors claim, are united by the human appreciation of the rainbow's compelling natural beauty. And who can disagree? An appendix is provided ("a field guide to the rainbow") comprising nineteen basic questions about the rainbow, with nontechnical (but scientifically accurate) answers for the interested observer. This is followed by a set of chapter notes and a bibliography, both of which are very comprehensive. The more technical scientific aspects of the distribution of light within a rainbow are scattered liberally throughout the latter half of the book; indeed, the reader more interested in the scientific aspects of the rainbow might wish to read the last five chapters in parallel with the first five (as did I). The technical aspects referred to are explained with great clarity.

But what *is* a rainbow? Towards the end of this review, several mutually inexclusive but complementary levels of explanation will be noted, reflecting the fact that there is a great deal of physics and mathematics behind one of nature's most awesome spectacles. At a more basic observational level, surely *everyone* can describe the colored arc of light we call a rainbow, we might suggest, certainly as far as the *primary* bow is concerned. In principle, whenever there is a primary rainbow, there is a larger and fainter *secondary* bow. The primary (formed by light being refracted twice and reflected once in raindrops) lies beneath a fainter secondary bow (formed by an additional reflection, and therefore fainter because of light loss) which is not always easily seen. There are several other things to look for: faint pastel fringes just below the top of the primary bow (supernumerary bows, of which more anon), the reversal of colors in the secondary bow, the dark region between the two bows, and the bright region below the primary bow. This observational description, however, is probably not one that is universally known; in some cultures it is considered unwise even to look at a rainbow. Indeed, in many parts of the world, it appears, merely pointing at a rainbow is considered to be a foolhardy act. Lee and Fraser state that "getting jaundice, losing an eye, being struck by lightning, or simply disappearing are among the unsavory aftermaths of rainbow pointing."

The historical descriptions are in places quite breathtaking; we are invited to look over the shoulders, as it were, of Descartes and Newton as they work through their respective accounts of the rainbow's position and colors. While the color theory of Descartes was flawed, his geometric theory was not. Commenting on the latter, the authors point out that "Descartes'[s] seventeenth-century analysis of the rainbow bears out Plato's

Figure 1. A sector illustrating the rainbow configuration, including both a primary bow (on the bottom) and a secondary bow (at the top). The bright region is not shown to scale.
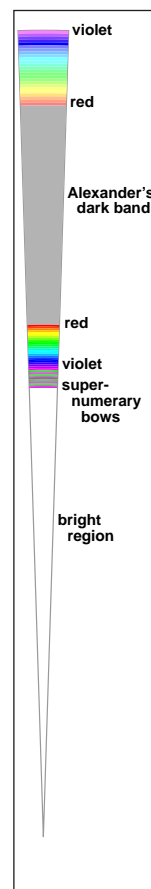
great faith in observations simplified and clarified by the power of mathematics." Newton, on the other hand, eschews Descartes's cumbersome ray-tracing technique and "silently invokes his mathematical invention of the 1660s, differential calculus, to specify the minimum deviation rays of the primary and secondary rainbows." Later in the book Lee and Fraser remark that Aristotle and later scientists in antiquity "constructed theories that primarily *describe* natural phenomena in mathematical or geometric terms, with little or no concern for physical mechanisms that might *explain* them." This contrast goes to the heart of the difference between Aristotelian and mathematical modeling.

I was pleasantly surprised to learn that the English painter John Constable was quite an avid amateur scientist: he was concerned that his paintings of clouds and rainbows should accurately reflect the science of the day, and he took great trouble to acquaint himself with Newton's theory of the rainbow (many other details can be found in pp. 80–7 of the book). In a similar vein, the writer John Ruskin was a detailed observer of nature, his goal being "that of transforming close observation into faithful depiction of a purposeful, divinely shaped nature." The poet John Keats implies in his *Lamia* that Newton's natural philosophy destroys the beauty of the rainbow ("There was an awful rainbow once in heaven: / We know her woof, her texture; she is given / In the dull catalog of common things. / Philosophy will clip an Angel's wings"). Keats's words reflected a continuing debate: did scientific knowledge facilitate or constrain poetic descriptions of nature? His contemporary William Wordsworth apparently held a different opinion, for he wrote, "The beauty in form of a plant or an animal is not made less but more apparent as a whole by more accurate insight into its constituent properties and powers."

In a subsection of Chapter 4 entitled "The Inescapable (and Unapproachable) Bow", the authors address, amongst other things, some common misperceptions about rainbows. Just as occurs when we contemplate our visage in a mirror, what we see is not an *object*, but an *image*; near the end of the chapter this rainbow image is beautifully portrayed as "a mosaic of sunlit rain." Optically, the rainbow is located at infinity, even
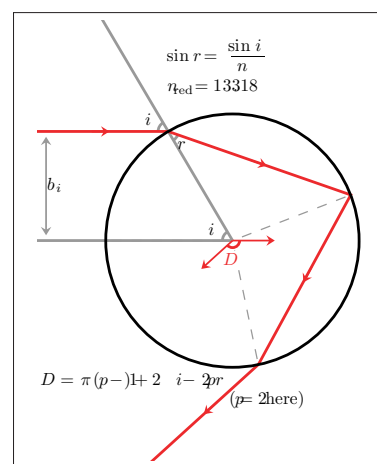

Figure 2. A typical path through a raindrop for a ray of light contributing to the primary rainbow, according to geometrical optics.
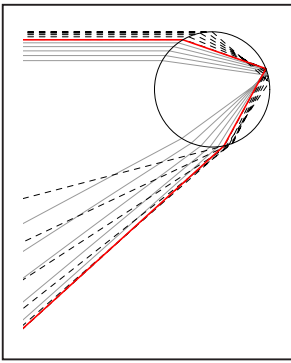
Figure 3. The paths of several rays of light impinging upon a raindrop with different angles of incidence. Such rays contribute to the primary rainbow. The ray of minimum deviation (the rainbow ray) is the darkest line.
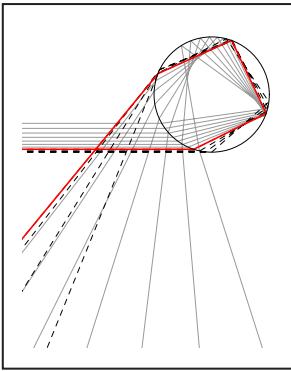

Figure 4. The ray paths for the secondary rainbow, similar to Figure 3 above. Again, the ray of minimum deviation is the darkest line.

though the raindrops or droplets sprayed from a garden hose are not. To place the rainbow in the sky, note that the *antisolar point* is 180° from the Sun on a line through the head of the observer. As will be noted below, rays of sunlight are deviated from this line (in a clockwise sense in Figure 2); the ray of minimum deviation (sometimes called the rainbow or Descartes ray) is deviated by about 138° for the primary bow. Therefore there is a concentration of deviated rays near this angle, and so for the observer the primary bow is an arc of a circle of radius 180° − 138° = 42° centered on the antisolar point. Thus the rainbow has a fixed angular radius of about 42° despite illusions (and allusions) to the contrary. The corresponding angle for the secondary bow is about 51°, though in each instance the angle varies a little depending on the wavelength of the light. Without this phenomenon of *dispersion* there would be only a "whitebow"!

It is perhaps an occupational hazard for professional scientists and mathematicians to be a little frustrated by incorrect depictions or explanations of observable phenomena and mathematical concepts in literature, art, the media, etc. From the point of view of an artist, however (Constable notwithstanding), scientific accuracy is not necessarily a prelude to artistic expression; indeed, to some it may be considered a hindrance. Nevertheless, we can identify perhaps with the authors, who, in commenting on the paintings of Frederic Church (in particular his *Rainy Season in the Tropics*), write, "Like Constable, he in places bends the unyielding rule that all shadows must be radii to the bow (that is, they must converge on the rainbow's center)." The same rule applies to sunbeams, should they be observable concurrently with a rainbow.

By the middle of the eighteenth century, the contributions of Descartes and Newton notwithstanding, observations of supernumerary bows were a persistent reminder of the inadequacy of current theories of the rainbow. As Lee and Fraser so pointedly remark, "One common reaction to being confronted with the unexplained is to label it inexplicable." This led to these troubling features being labeled spurious; hence the unfortunate addition of the adjective *supernumerary* to the rainbow phenomenon. Now it is known that such bows "are an integral part of the rainbow, not a vexing corruption of it"; an appropriate, though possibly unintended, mathematical pun (see below). By focusing attention on the light *wavefronts* incident on a spherical drop rather than on the rays normal to them, it is easier to appreciate the self-interference of such a wave as it becomes "folded" onto itself as a result of refraction and reflection within the drop. This is readily seen from Figures 8.7 and 8.9 in the book (see Figures 5 and 6 here), from which the true extent of the rainbow is revealed: the primary rainbow is in fact the *first interference maximum*, the second and third maxima being the first and second supernumerary bows respectively (and so on).

The angular spacing of these bands depends on the size of the droplets producing them. The width of individual bands and the spacing between them decreases as the drops get larger. If drops of many different sizes are present, these supernumerary arcs tend to overlap somewhat and smear out what would have been obvious interference bands for droplets of uniform size. This is why these pale blue or pink or green bands are then most noticeable near the top of the rainbow: it is the near-sphericity of the smaller drops that enable them to contribute to this part of the bow; larger drops are distorted from sphericity by the aerodynamic forces acting upon them. Nearer the horizon a wide range of drop size contributes to the bow, but at the same time it tends to blur the interference bands. In principle, similar interference effects also occur above the secondary rainbow, though they are very rare, for reasons discussed in the penultimate chapter. Lee and Fraser summarize the importance of spurious bows with typical metaphorical creativity: "Thus the supernumerary rainbows proved to be the

$y = Ai(x)$
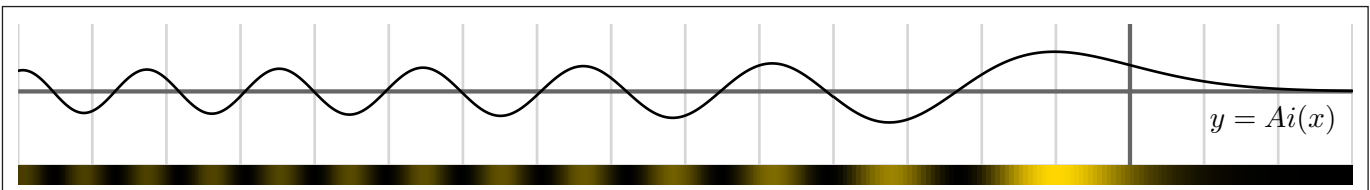
Graph of the Airy function $Ai(x)$, which is a solution of the differential equation $y'' + xy = 0$. Below, according to Airy's theory, illumination is proportional to $Ai(x)^2$, simulating a monochromatic bow along with several supernumerary bows.
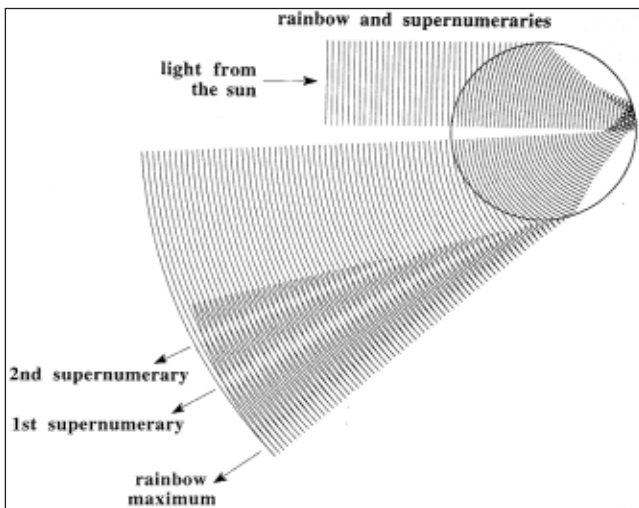
Figure 5. A wavefront version of Figure 3 illustrating the constructive and destructive interference patterns of the light as it interacts with a raindrop. The width of each band of color is relatively narrow, resulting in fairly pure rainbow colors. The primary rainbow and the first two supernumerary bows are identified here.
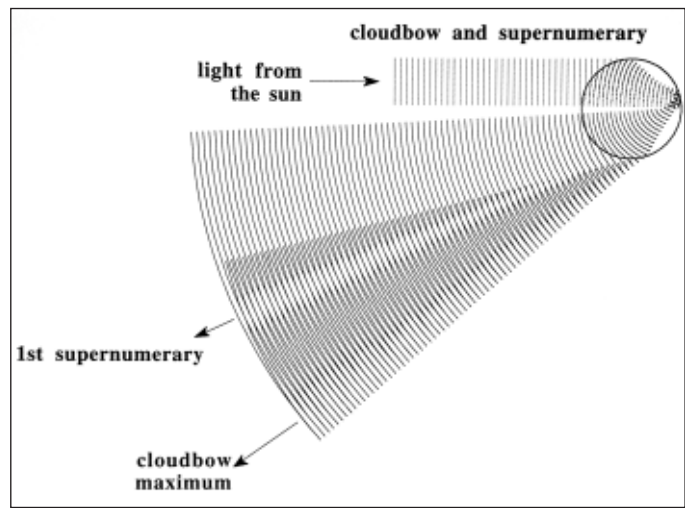

Figure 6. Similar to Figure 5, but for a much smaller cloud droplet. The resulting bands of color are so broad that additive color mixing produces a whitish "cloudbow".

midwife that delivered the wave theory of light to its place of dominance in the nineteenth century."

It is important to recognize that, not only were the Cartesian and Newtonian theories unable to account for the presence of supernumerary bows, but also they both predicted an *abrupt* transition between regions of illumination and shadow (as at the edges of Alexander's dark band, when only rays giving rise to the primary and secondary bows are considered). In the wave theory of light such sharp boundaries are softened by *diffraction*, which occurs when the normal interference pattern responsible for rectilinear propagation of light is distorted in some way. Diffraction effects are particularly prevalent in the vicinity of *caustics*. In 1835 Potter showed that the rainbow ray may be interpreted as a caustic, i.e., the envelope of the system of rays constituting the rainbow. The word *caustic* means burning, and caustics are associated with regions of high-intensity illumination (with geometrical optics predicting an infinite intensity there). Thus the rainbow problem is essentially that of determining the intensity of (scattered) light in the neighborhood of a caustic.

This was exactly what Airy attempted to do several years later in 1838. The principle behind Airy's approach was established by Huygens in the seventeenth century: Huygens's principle regards every point of a wavefront as a secondary source of waves, which in turn defines a new wavefront and hence determines the subsequent propagation of the wave. As pointed out by Nussenzveig [9], Airy reasoned that if one knew the amplitude distribution of the waves along any complete wavefront in a raindrop, the distribution at any other point could

be determined by Huygens's principle. Airy chose as his starting point a wavefront surface inside the raindrop, the surface being orthogonal to all the rays that constitute the primary bow; this surface has a point of inflection wherever it intersects the ray of minimum deviation, the "rainbow ray". Using the standard assumptions of diffraction theory, he formulated the local intensity of scattered light in terms of a "rainbow integral", subsequently renamed the *Airy integral* in his honor; it is related to the now familiar Airy function. It is analogous to the Fresnel integrals which also arise in diffraction theory. There are several equivalent representations of the Airy integral in the literature; following the form used by Nussenzveig [10], we write it as

$$(1) \quad \text{Ai}(C) = (3^{1/3}\pi^{-1}) \int_0^\infty \cos(t^3 + 3^{1/3}Ct)\,dt.$$

While the argument of the Airy function is arbitrary at this point, C refers to the set of control space parameters in the discussion below on diffraction catastrophes. In this case it represents the deviation or scattering angle coordinate. One severe limitation of the Airy theory for the optical rainbow is that the amplitude distribution along the initial wavefront is unknown: based on certain assumptions it has to be guessed. There is a natural and fundamental parameter, the *size parameter*, β, which is useful in determining the domain of validity of the Airy approximation; it is defined as the ratio of the droplet circumference to the wavelength η of light. In terms of the wavenumber k this is

$$\beta = \frac{2\pi a}{\eta} \equiv ka,$$

a being the droplet radius. Typically, for sizes ranging from fog droplets to large raindrops, β ranges from about one hundred to several thousand. Airy's approximation is a good one only for $\beta \gtrsim 5000$ and for angles sufficiently close to that of the rainbow ray. In light of these remarks it is perhaps surprising that an exact solution does exist for the rainbow problem, as indicated below.

In the epilogue to the final chapter of the book, one subsection is entitled "Airy's Rainbow Theory: The Incomplete 'Complete' Answer". This theory did go beyond the models of the day in that it quantified the dependence upon the raindrop size of (i) the rainbow's angular width, (ii) its angular radius, and (iii) the spacing of the supernumeraries. Also, unlike the models of Descartes and Newton, Airy's predicted a nonzero distribution of light intensity in Alexander's dark band and a finite intensity at the angle of minimum deviation (as noted above, the earlier theories predicted an infinite intensity there). However, spurred on by Maxwell's recognition that light is part of the electromagnetic spectrum and the subsequent publication of his mathematical treatise on electromagnetic waves, several mathematical physicists sought a more complete theory of scattering, because it had been demonstrated by then that the Airy theory failed to predict precisely the angular position of many laboratory-generated rainbows. Among them were the German physicist Gustav Mie,[1] who published a paper in 1908 on the scattering of light by homogeneous spheres in a homogeneous medium, and Peter Debye, who independently developed a similar theory for the scattering of electromagnetic waves by spheres. Mie's theory was intended to explain the colors exhibited by colloidally dispersed metal particles, whereas Debye's work, based on his 1908 thesis, dealt with the problem of light pressure on a spherical particle. The resulting body of knowledge is usually referred to as Mie theory, and typical computations based on it are formidable compared with those based on Airy theory, unless the drop size is sufficiently small. A similar (but nonelectromagnetic) formulation arises

---

[1]*I am grateful to J. D. Jackson, who informed me that "Ludvig Lorenz, a Danish theorist, preceded Mie by about fifteen years in the treatment of the scattering of electromagnetic waves by spheres." His contributions to electromagnetic scattering theory and optics are rather overlooked, probably because his work was published in Danish (in 1890). Further details of his research, including his contributions to applied mathematics, may be found in reference [5] (the article immediately following that one on p. 4696 is about Gustav Mie). There is also a valuable historical account by Logan [6] of other contributions to this branch of (classical) mathematical physics.*

in the scattering of sound waves by an impenetrable sphere, studied by Lord Rayleigh and others in the nineteenth century [6].

Mie theory is based on the solution of Maxwell's equations of electromagnetic theory for a monochromatic plane wave from infinity incident upon a homogeneous isotropic sphere of radius a. The surrounding medium is transparent (as the sphere may be), homogeneous, and isotropic. The incident wave induces forced oscillations of both free and bound charges in synchrony with the applied field, and this induces a secondary electric and magnetic field, each of which has components inside and outside the sphere. Of crucial importance in the theory are the scattering amplitudes ($S_j(k, \theta)$, $j = 1, 2$) for the two independent polarizations, θ being the angular variable; these amplitudes can be expressed as an infinite sum called a partial-wave expansion. Each term (or partial wave) in the expansion is defined in terms of combinations of Legendre functions of the first kind, Riccati-Bessel functions, and Riccati-Hankel functions (the latter two being rather simply related to spherical Bessel and Hankel functions respectively).

It is obviously of interest to determine under what conditions such an infinite set of terms can be truncated and what the resulting error may be by so doing. However, it turns out that the number of terms that must be retained is of the same order of magnitude as the size parameter β, i.e., up to several thousand for the rainbow problem. On the other hand, the "why is the sky blue?" scattering problem—*Rayleigh scattering*—requires only one term, because the scatterers are molecules much smaller than a wavelength of light, so the simplest truncation—retaining only the first term—is perfectly adequate. Although in principle the rainbow problem can be "solved" with enough computer time and resources, numerical solutions by themselves (as Nussenzveig points out [9]) offer little or no insight into the physics of the phenomenon.

The *Watson transform*, originally introduced by Watson in connection with the diffraction of radio waves around the Earth (and subsequently modified by Nussenzveig in his studies of the rainbow problem), is a method for transforming the slowly converging partial-wave series into a rapidly convergent expression involving an integral in the complex angular-momentum plane. The Watson transform is intimately related to the *Poisson summation formula*

$$(2) \quad \sum_{l=0}^{\infty} g(l + \frac{1}{2}, x) = \sum_{m=-\infty}^{\infty} e^{-im\pi} \int_0^{\infty} g(\lambda, x) e^{2\pi i m \lambda} \, d\lambda$$

for an "interpolating function" $g(\lambda, x)$, where x denotes a set of parameters and $\lambda = l + \frac{1}{2}$ is now considered to be the complex angular momentum variable.

But why *angular momentum*? Although they possess zero rest mass, photons have energy $E = hc/\eta$ and momentum $E/c = h/\eta$, where h is Planck's constant and c is the speed of light in vacuo. Thus for a nonzero impact parameter b_i, a photon will carry an angular momentum $b_i h/\eta$ (b_i being the perpendicular distance of the incident ray from the axis of symmetry of the sun-raindrop system). Each of these discrete values can be identified with a term in the partial-wave series expansion. Furthermore, as the photon undergoes repeated internal reflections, it can be thought of as orbiting the center of the raindrop. Why *complex* angular momentum? This allows the above transformation to effectively "redistribute" the contributions to the partial wave series into a few points in the complex plane—specifically poles (called *Regge* poles in elementary particle physics) and saddle points. Such a decomposition means that angular momentum, instead of being identified with certain discrete real numbers, is now permitted to move continuously through complex values. However, despite this modification, the poles and saddle points have profound physical interpretations in the rainbow problem.

In the simplest Cartesian terms, on the illuminated side of the rainbow (in a limiting sense) there are two rays of light emerging in parallel directions: at the rainbow angle they coalesce into the ray of minimum deflection, and on the shadow side, according to geometrical optics, they vanish (this is actually a good definition of a caustic curve or surface). From a study of real and complex rays, it happens that, mathematically, in the context of the complex angular momentum plane, a rainbow is the *collision of two real saddle points*. But this is not all: this collision does not result in the mutual annihilation of these saddle points; instead, two complex saddle points are born, one corresponding to a complex ray on the shadow side of the caustic curve. This is directly associated with the diffracted light in Alexander's dark band.

Lee and Fraser point out that as far as most aspects of the optical rainbow are concerned, Mie theory is esoteric overkill; Airy theory is quite sufficient for describing the outdoor rainbow. I found particularly valuable the following comment by the authors; it has implications for mathematical modeling in general, not just for the optics of the rainbow. In their comparison of the less accurate Airy theory of the rainbow with the more general and powerful Mie theory, they write, "Our point here is not that the exact Mie theory describes the natural rainbow inadequately, but rather that the approximate Airy theory can describe it quite well. Thus the supposedly outmoded Airy theory generates a more natural-looking map of real rainbow colors than Mie theory does, even though Airy theory makes substantial errors in describing the scattering of monochromatic light by isolated small drops. *As in many hierarchies of scientific models, the virtues of a simpler theory can, under the right circumstances, outweigh its vices*" [emphasis added].

Earlier in this review the question, What is a rainbow? was answered at a basic descriptive level. But at an *explanatory* level a rainbow is much, much more than such an answer might imply. Amongst other things a rainbow is: (1) a concentration of light rays corresponding to a minimum (for the primary bow) of the deviation or scattering angle as a function of the angle of incidence; this minimum is identified as the Descartes or rainbow ray; (2) a caustic, separating a 2-ray region from a 0-ray (or shadow) region; (3) an integral superposition of waves over a (locally) cubic wavefront (the Airy approximation); (4) an interference problem (the origin of the supernumerary bows); (5) a coalescence of two real saddle points; (6) a result of scattering by an effective potential consisting of a square well and a centrifugal barrier (a surprising macroscopic quantum connection); (7) associated with tunneling in the so-called *edge domain* [4]; (8) a tangentially polarized circular arc; and (9) a *fold diffraction catastrophe*. These nine topics are very much interrelated, and the interested reader will find many references to them in [1], [4], [9], and [10]. The books by Grandy and Nussenzveig are excellent: they are primarily written by theoretical physicists for theoretical physicists and so do not possess the kind of rigor analysts seek (such as in the book by Pearson [11]); perhaps such a transition will be made in the future. But why not read *The Rainbow Bridge* first?

A Mathematical Appendix

Some descriptive material has been presented above concerning complementary explanations of the rainbow. It is of interest to put some mathematical flesh on those bones, so to that end a brief summary is provided of three aspects of the rainbow problem at the level of geometrical optics, as a scalar scattering problem, and as a diffraction catastrophe. These accounts touch on items (1)-(3), (5), (6) and (9) in particular, but as noted above it is difficult to separate them in a precise way.

A rainbow occurs when the the scattering angle D, as a function of the angle of incidence i, passes through an extremum (a minimum for the primary bow; see Figure 3). The "folding back" of the corresponding scattered or deviated ray takes place at this extremal scattering angle (the rainbow angle $D_{\min} \equiv \theta_R$; note that sometimes in the literature the rainbow angle is defined as the complement of the deviation, $\pi - D_{\min}$). Two rays scattered in the same direction with different angles of incidence on the illuminated side of the rainbow ($\theta > \theta_R$) fuse together at the rainbow angle and disappear

as the dark side ($\theta < \theta_R$) is approached. This is one of the simplest physical examples of a *fold catastrophe* in the sense of Thom. As is noted below, rainbows of different orders are associated with so-called *Debye terms* of different orders; the primary and secondary bows correspond to $p = 2$ and $p = 3$ respectively. For $p - 1$ internal reflections, p being an integer greater than 1, the total deviation is, from elementary geometry,

$$D_{p-1}(i) = 2(i - pr) + (p - 1)\pi,$$

where by Snell's law the angle of refraction $r = r(i)$ is also a function of the angle of incidence. Thus for the primary rainbow ($p = 2$),

$$D_1(i) = \pi - 4r + 2i,$$

and for the secondary rainbow ($p = 3$),

$$D_2(i) = 2i - 6r$$

(modulo 2π). Removing the dependence on r, the angle through which a ray is deviated for $p = 2$ is

$$D_1(i) = \pi + 2i - 4\arcsin\left(\frac{\sin i}{n}\right),$$

where $n > 1$ is the refractive index of the raindrop. In general

$$D_{p-1}(i) = (p - 1)\pi + 2i - 2p\arcsin\left(\frac{\sin i}{n}\right).$$

The minimum deviation occurs for $p = 2$ when

$$i = i_c = \arccos\sqrt{\frac{n^2 - 1}{3}}$$

and in general when

$$i = i_c = \arccos\sqrt{\frac{n^2 - 1}{p^2 - 1}}.$$

For the primary rainbow $p = 2$, so $i_c \approx 59°$, using an approximate value for water of $n = 4/3$; from this it follows that $D(i_c) = D_{\min} = \theta_R \approx 138°$. For the secondary bow $p = 3$ and $i_c \approx 72°$; now $D(i_c) = \theta_R \approx 231° = -129°$. The rainbows each lie on the surface of a cone, centered at the observer's eye, with axis the line from that point to the antisolar point (delineated by the shadow of the observer's head). The cone semiangles are about $42°$ and $51°$ respectively for the primary and secondary bows. For $p = 2$, in terms of n alone, $D(i_c) = \theta_R$ (the rainbow angle) is defined by

$$D(i_c) = \theta_R = 2\arccos\left[\frac{1}{n^2}\left(\frac{4 - n^2}{3}\right)^{3/2}\right].$$

This approach can be thought of as the elementary classical description. Surprising as it may seem, there is also a wave mechanical counterpart to the optical rainbow; indeed, such effects are well known in atomic, molecular, and nuclear physics [1], [10]. On the way to this, so to speak, there is a "semiclassical" description. In a primitive sense, the semiclassical approach is the "geometric mean" between classical and quantum mechanical descriptions of phenomena in which interference and diffraction effects enter the picture. The latter do so via the transition from geometrical optics to wave optics. A measure of the scattering (by raindrop or atom) that occurs is the differential scattering cross-section $d\sigma/d\Omega = |f(k, \theta)|^2$ (essentially the relative particle or wave "density" scattered per unit solid angle at a given angle θ in cylindrically symmetric geometry), where $f(k, \theta)$ is the *scattering amplitude* defined below; this in turn is expressible as a partial-wave expansion. The formal relationship between the latter and the classical differential cross-section is established using the WKB approximation, and the principle of stationary phase is used to evaluate asymptotically a certain phase integral. A point of stationary phase can be identified with a classical trajectory, but if more than one such point is present (provided they are well separated and of the first order), the corresponding expression for $|f(k, \theta)|^2$ will contain interference terms. This is a distinguishing feature of the "primitive" semiclassical formulation. Indeed, the infinite intensities predicted by geometrical optics at focal points, lines, and caustics in general are breeding grounds for diffraction effects, as are light/shadow boundaries for which geometrical optics predicts finite discontinuities in intensity. Such effects are most significant when the wavelength is comparable with (or larger than) the typical length scale for variation of the physical property of interest (e.g., size of the scattering object). Thus a scattering object with a "sharp" boundary (relative to one wavelength) can give rise to *diffractive scattering* phenomena [10].

There are critical angular regions where this semiclassical approximation breaks down and diffraction effects cannot be ignored, although the angular ranges in which such critical effects become significant get narrower as the wavelength decreases. Early work in this field contained *transitional asymptotic approximations* to the scattering amplitude in these critical angular domains, but they have very narrow domains of validity and do not match smoothly with neighboring noncritical angular domains. It is therefore of considerable importance to seek *uniform asymptotic approximations* that by definition do not suffer from these failings. Fortunately, the problem of plane-wave scattering by a homogeneous sphere exhibits all of the critical scattering effects (and it can be solved exactly, in principle) and is therefore an ideal laboratory in which to test both the efficacy and accuracy of the various approximations. Furthermore, it has

relevance to both quantum mechanics (as a square well or barrier problem) and optics (Mie scattering); indeed, it also serves as a model for the scattering of acoustic and elastic waves and, as noted earlier, was studied in the early twentieth century as a model for the diffraction of radio waves around the surface of the earth. Light waves obviously are electromagnetic in character, and so the full weight of that theory is necessary to explain features such as the polarization of the rainbow, but the essence of the approach is well captured by the scalar theory.

The essential mathematical problem for scalar waves can be thought of either in terms of classical mathematical physics, e.g., the scattering of sound waves, or in wave-mechanical terms, e.g., the nonrelativistic scattering of particles by a square potential well (or barrier) of radius a and depth (or height) V_0. In either case we can consider a scalar plane wave impinging in the direction $\theta = 0$ on a penetrable ("transparent") sphere of radius a. The wave function $\psi(r, \theta)$ satisfies the scalar Helmholtz equation

$$\begin{aligned} \nabla^2 \psi + n^2 k^2 \psi = 0 & \quad r \leq a \\ \nabla^2 \psi + k^2 \psi = 0 & \quad r \geq a, \end{aligned}$$

(3)

where again k is the wavenumber and $n > 1$ is the refractive index of the sphere (a similar problem can be posed for gas bubbles in a liquid, for which $n < 1$). The boundary conditions are that $\psi(r, \theta)$ and $\psi'(r, \theta)$ are continuous at the surface. Furthermore, at large distances from the sphere ($r \gg a$) the wave field can be decomposed into an incident wave + scattered field, i.e.,

$$\psi \sim e^{ikr \cos \theta} + \frac{af(k, \theta)e^{ikr}}{r},$$

the dimensionless scattering amplitude being defined as

$$(4) \quad f(k, \theta) = \frac{1}{2ika} \sum_{l=0}^{\infty} (2l + 1)(S_l(k) - 1)P_l(\cos \theta),$$

where S_l is the scattering function for a given l and P_l is a Legendre polynomial of degree l. For a spherical square well or barrier,

$$S_l = -\frac{h_l^{(2)}(\beta)}{h_l^{(1)}(\beta)} \left\{ \frac{\ln' h_l^{(2)}(\beta) - n \ln' j_l(\alpha)}{\ln' h_l^{(1)}(\beta) - n \ln' j_l(\alpha)} \right\},$$

where $\ln'$ represents the logarithmic derivative operator, j_l and h_l are spherical Bessel and Hankel functions respectively, the size parameter $\beta = ka$ plays the role of a dimensionless external wavenumber, and $\alpha = n\beta$ is the corresponding *internal* wavenumber. S_l may be equivalently expressed in terms of cylindrical Bessel and Hankel functions. The lth partial wave in the solution is associated
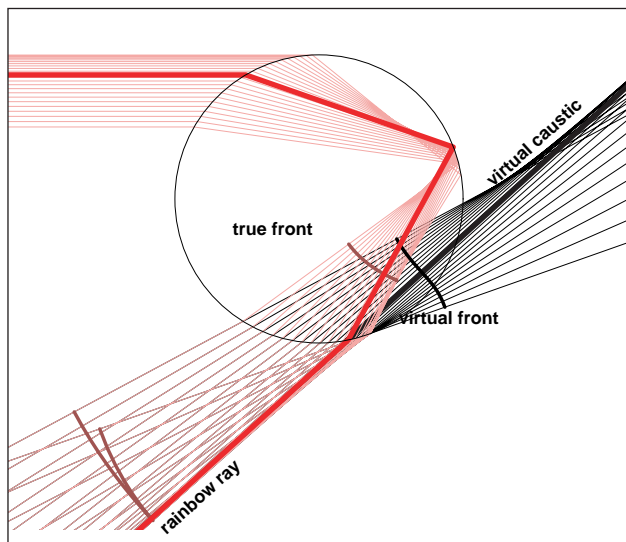

Figure 7. Wavefronts in a raindrop. The caustic is best seen in the virtual rays (projected back from the final exterior rays), as is the basis of Airy's analysis. After the virtual (or Airy) wavefront emerges, the part which is convex forward continues to expand, while the concave forward part collapses to a focus and then expands. This is the reason for the cusped wavefront at the bottom left of the figure.

with an impact parameter $b_l = (l + \frac{1}{2})/k$; i.e., only rays hitting the sphere ($b_l \lesssim a$) are significantly scattered, and the number of terms that must be retained in the series to get an accurate result is of order β. As implied earlier, for visible light scattered by water droplets in the atmosphere, $\beta \sim$ several thousand. This is why, to quote Arnold Sommerfeld, "The series converge so slowly that they become practically useless." This problem can be remedied by using the Poisson summation formula (2) above to rewrite $f(k, \theta)$ as

$$\begin{aligned} f(\beta, \theta) = \frac{i}{\beta} \sum_{m=-\infty}^{\infty} (-1)^m \\ \times \int_0^{\infty} [1 - S(\lambda, \beta)]P_{\lambda - \frac{1}{2}}(\cos \theta)e^{2im\pi\lambda} \lambda \, d\lambda. \end{aligned}$$

(5)

For fixed β, $S(\lambda, \beta)$ is a meromorphic function of the complex variable $\lambda = l + 1/2$, and in particular in what follows it is the *poles* of this function that are of interest. In terms of cylindrical Bessel and Hankel functions, the poles are defined by the condition

$$\ln' H_{\lambda}^{(1)}(\beta) = n \ln' J_{\lambda}(\alpha)$$

and are called *Regge poles* in the scattering theory literature [7], [8]. Typically, they are associated with surface waves for the *impenetrable* sphere problem, but for the *transparent* sphere *two* types of Regge poles arise—one type (*Regge-Debye poles*) leading to rapidly convergent residue series, representing the surface wave (or diffracted or creeping ray) contributions to the scattering amplitude; and the other

type associated with resonances via the internal structure of the potential, which is now of course accessible. They are characterized by an effective radial wavenumber within the potential well. Many in this latter group are clustered close to the real axis, spoiling the rapid convergence of the residue series. They correspond to different types of resonances depending on the value of $\text{Re}(\lambda_j)$. Thus mathematically these resonances are complex eigenfrequencies associated with the poles λ_j of the scattering function $S(\lambda, k)$ in the first quadrant of the complex λ-plane for real values of β. The imaginary parts of the poles are directly related to resonance widths (and therefore lifetimes). As β increases, the poles λ_j trace out Regge trajectories and $\text{Im}\,\lambda_j$ tends exponentially to zero. When $\text{Re}\,\lambda_j$ passes close to a "physical" value, $\lambda = l + 1/2$, it is associated with a resonance in the lth partial wave; the larger the value of β, the sharper the resonance becomes for a given node number j. The other type of significant points to be considered in the complex λ-plane are saddle points (which can be real or complex). In the complex angular momentum method, contour integral paths are deformed to concentrate the dominant contributions to the poles and saddle points. A real saddle point is also a point of stationary phase, and stationarity implies that Fermat's principle is satisfied. Since the phase in many cases at such points can be approximated by the WKB phase, the latter can be identified with the rays of geometrical optics (sometimes called classical paths). The asymptotic contributions from these points are the dominant ones within geometrically illuminated regions [10].

In [8] it is shown that

$$
\begin{aligned}
S(\lambda, \beta) = {} & \frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)} R_{22}(\lambda, \beta) \\
(6) \quad & + T_{21}(\lambda, \beta) T_{12}(\lambda, \beta) \frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)} \sum_{p=1}^{\infty} [\rho(\lambda, \beta)]^{p-1},
\end{aligned}
$$

where

$$
\rho(\lambda, \beta) = R_{11}(\lambda, \beta) \frac{H_\lambda^{(1)}(\alpha)}{H_\lambda^{(2)}(\alpha)}.
$$

This is the *Debye expansion*, arrived at by expanding the expression $[1 - \rho(\lambda, \beta)]^{-1}$ as an infinite geometric series. R_{22}, R_{11}, T_{21}, and T_{12} are respectively the external/internal reflection and internal/external transmission coefficients for the problem. This procedure transforms the interaction of "wave + sphere" into a series of surface interactions. In so doing it unfolds the stationary points of the integrand so that a given integral in the Poisson summation contains at most one stationary point. This permits a ready identification of the many terms in accordance with ray theory. The first term represents direct reflection from the surface. The pth

term in the summation represents transmission into the sphere (via the term T_{21}) subsequently bouncing back and forth between $r = a$ and $r = 0$ a total of p times with $p - 1$ internal reflections at the surface (this time via the R_{11} term in ρ). The final factor in the second term, T_{12}, corresponds to transmission to the outside medium. In general, therefore, the pth term of the Debye expansion represents the effect of $p + 1$ surface interactions. Now $f(\beta, \theta)$ can be expressed as

$$
(7) \qquad f(\beta, \theta) = f_0(\beta, \theta) + \sum_{p=1}^{\infty} f_p(\beta, \theta),
$$

where

$$
\begin{aligned}
(8) \quad f_0(\beta, \theta) = {} & \frac{i}{\beta} \sum_{m=-\infty}^{\infty} (-1)^m \int_0^\infty \left(1 - \frac{H_\lambda^{(2)}(\beta)}{H_\lambda^{(1)}(\beta)} R_{22} \right) \\
& \times P_{\lambda - \frac{1}{2}}(\cos \theta) \exp(2im\pi\lambda) \lambda \, d\lambda.
\end{aligned}
$$

The expression for $f_p(\beta, \theta)$ involves a similar type of integral for $p \geq 1$. The application of the modified Watson transform to the third term ($p = 2$) in the Debye expansion of the scattering amplitude shows that it is this term which is associated with the phenomena of the primary rainbow. Residue contributions also arise from the Regge-Debye poles. More generally, for a Debye term of given order p, a rainbow is characterized in the λ-plane by the occurrence of two real saddle points, $\bar{\lambda}$ and $\bar{\lambda}'$, between 0 and β in some domain of scattering angles θ, corresponding to the two scattered rays on the lighted side. As $\theta \to \theta_R^+$ the two saddle points move toward each other along the real axis, merging together at $\theta = \theta_R$. As θ moves into the dark side, the two saddle points become complex, moving away from the real axis in complex conjugate directions. Therefore, from a mathematical point of view, *a rainbow can be defined as a coalescence of two saddle points in the complex angular momentum plane.*

The rainbow light/shadow transition region is thus associated physically with the confluence of a pair of geometrical rays and their transformation into complex rays; mathematically this corresponds to a pair of real saddle points merging into a complex saddle point. The next problem is to find the asymptotic expansion of an integral having two saddle points that move toward or away from each other. The generalization of the standard saddle-point technique to include such problems was made by Chester et al. [3]. Using their method, Nussenzveig was able to find a uniform asymptotic expansion of the scattering amplitude which was valid throughout the rainbow region and which matched smoothly onto results for neighboring regions [10], [8]. Unsurprisingly perhaps, the lowest-order approximation in this expansion turns out

to be the *Airy approximation*, which, as already noted, was the best prior approximate treatment. However, Airy's theory has a limited range of applicability as a result of its underlying assumptions, e.g., $\beta \gtrsim 5000$, $|\theta - \theta_R| \lesssim 0.5°$. By contrast, the uniform expansion (and more generally the complex angular momentum theory) is valid over much larger ranges.

Finally, we examine the rainbow as a diffraction catastrophe following the account by Berry and Upstill [2]. Optics is concerned to a great degree with families of rays filling regions of space; the *singularities* of such ray families are *caustics*. For optical purposes this level of description is important for classifying caustics using the concept of *structural stability*: this enables one to classify those caustics whose topology survives perturbation. Structural stability means that if a singularity S_1 is produced by a generating function ϕ_1, and ϕ_1 is perturbed to ϕ_2, the correspondingly changed S_2 is related to S_1 by a *diffeomorphism* of the control set C (that is, by a smooth reversible set of control parameters, a smooth deformation). In the present context this means physically that distortions of the raindrop shape to incoming wavefronts from their "ideal" spherical or planar forms do not prevent the formation of rainbows, though there may be some changes in the features. Another way of expressing this concept is to describe the system as well posed in the limited sense that small changes in the input generate correspondingly small changes in the output. For the elementary catastrophes, structural stability is a generic property of caustics. Each structurally stable caustic has a characteristic diffraction pattern, the wave function of which has an integral representation in terms of the standard polynomial describing that catastrophe. From a mathematical point of view, these diffraction catastrophes are especially interesting, because they constitute a new hierarchy of functions distinct from the special functions of analysis [2].

As noted above, at the level of geometrical optics the scattering deviation angle D has an extremum corresponding to the rainbow angle (or Descartes ray) when considered as a function of the angle of incidence i. This extremum is a minimum for the primary rainbow. Clearly the point $(i, D(i))$ corresponding to this minimum is a singular point (approximately $(59°, 138°)$) insofar as it separates a two-ray region ($D > D_{min} = \theta_R$) from a zero-ray region ($D < \theta_R$) at this level of description. This is a singularity or *caustic point*. The rays form a directional caustic at this point; this is a *fold catastrophe*, the simplest example of a catastrophe. It is the only stable singularity with *codimension* one (the dimensionality of the control space (one) minus the dimensionality of the singularity itself (zero)). In

physical space the caustic surface is asymptotic to a cone with semiangle $42°$.

Diffraction is discussed in terms of the scalar Helmholtz equation (3) at the point $\mathcal{R}$ for the complex scalar wavefunction $\psi(\mathcal{R})$. The concern in catastrophe optics is to study the asymptotic behavior of wave fields near caustics in the short-wave limit $k \to \infty$ (semiclassical theory). In a standard manner, $\psi(\mathcal{R})$ is expressed as

$$\psi(\mathcal{R}) = A(\mathcal{R})e^{i\kappa\chi(\mathcal{R})},$$

where the modulus A and the phase $\kappa\chi$ are both real quantities. The integral representation for ψ is

(9)
$$\psi(\mathcal{R}) = e^{-iN\pi/4} \left(\frac{\kappa}{2\pi} \right)^{N/2}$$
$$\times \int \cdots \int b(s; \mathcal{R}) \exp[ik\phi(s; \mathcal{R})] \, d^N s,$$

where N is the number of state (or behavior) variables s and b is a weight function. According to the principle of stationary phase, the main contributions to the above integral for given $\mathcal{R}$ come from the stationary points, i.e., those points s_i for which the gradient map $\partial\phi/\partial s_i$ vanishes; caustics are *singularities* of this map, where two or more stationary points coalesce. Because $k \to \infty$, the integrand is a rapidly oscillating function of s, so other than near the points s_i, destructive interference occurs and the corresponding contributions are negligible. The stationary points are well separated provided $\mathcal{R}$ is not near a caustic; the simplest form of stationary phase can then be applied and yields a series of terms of the form

$$\psi(\mathcal{R}) \approx \sum_\mu A_\mu \exp[ig_\mu(k, \mathcal{R})],$$

where the details of the g_μ need not concern us here. Near a caustic, however, two or more of the stationary points are close (in some appropriate sense), and their contributions cannot be separated without a reformulation of the stationary phase principle to accommodate this or by using diffraction catastrophes. The problem is that the ray contributions can no longer be considered separately; when the stationary points approach closer than a distance $\mathcal{O}(k^{-1/2})$, the contributions are not separated by a region in which destructive interference occurs. When such points coalesce, $\phi(s; \mathcal{R})$ is stationary to higher than first order and quadratic terms as well as linear terms in $s - s_\mu$ vanish. This implies the existence of a set of displacements ds_i, away from the extrema s_μ, for which the gradient map $\partial\phi/\partial s_i$ still vanishes, i.e., for which

$$\sum_i \frac{\partial^2\phi}{\partial s_i \partial s_j} ds_i = 0.$$
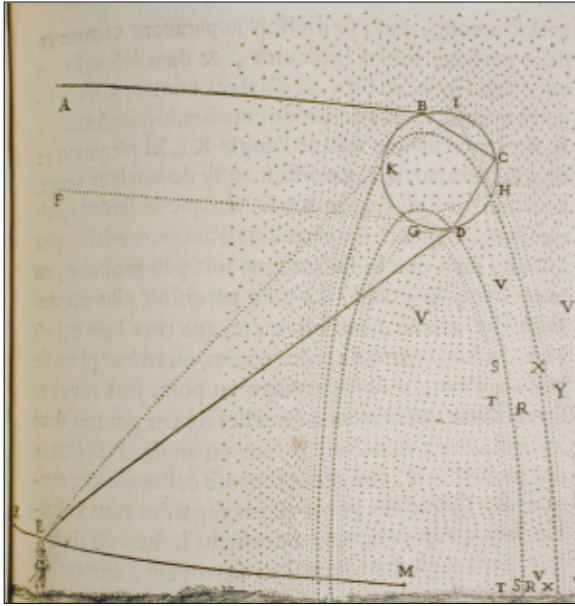
René Descartes correctly explained the reasons for the primary and secondary rainbows within the confines of geometrical optics. This image is from a copy of the first edition of "Les météores" at the Thomas Fisher Rare Book Library of the University of Toronto.

The condition for this homogeneous system of equations to have a solution (i.e., for the set of control parameters C to lie on a caustic) is that the Hessian

$$H(\phi) \equiv \det(\frac{\partial^2 \phi}{\partial s_i \partial s_j}) = 0$$

at points $s_\mu(C)$ where $\partial\phi/\partial s_i = 0$. The caustic defined by $H = 0$ determines the bifurcation set for which at least two stationary points coalesce (in the present circumstance this is just the rainbow angle). In view of this discussion there are two other ways of expressing this: (i) rays coalesce on caustics, and (ii) caustics correspond to singularities of gradient maps.

To remedy this problem, the function ϕ is replaced by a simpler "normal form" Φ with the same stationary-point structure, and the resulting diffraction integral is evaluated exactly. This is where the property of structural stability is so important, because if the caustic is structurally stable, it must be equivalent to one of the catastrophes (in the diffeomorphic sense described above). The result is a generic diffraction integral which will occur in many different contexts. The basic diffraction catastrophe integrals (one for each catastrophe) may be reduced to the form

$$(10) \quad \Psi(C) = \frac{1}{(2\pi)^{N/2}} \int \cdots \int \exp[i\Phi(s; C)] \, d^N s,$$

where s represents the state variables and C the control parameters (for the case of the rainbow

there is only one of each, so $N = 1$). These integrals stably represent the wave patterns near caustics. For the fold,

$$(11) \qquad \Phi(s; C) = \frac{1}{3}s^3 + Cs.$$

The diffraction catastrophes $\Psi(C)$ provide transitional approximations, valid close to the caustic and for short waves but increasingly inaccurate far from the caustic. By substituting the cubic term (11) into the integral (10), we obtain

$$(12) \qquad \Psi(C) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp[i(s^3/3 + Cs)] \, ds,$$

which is closely related to the Airy integral $\mathrm{Ai}(C)$ (equation (1), with $s = 3^{1/3}t$).

Let us now end where we began: at the rainbow bridge, but this time accompanied by a quotation from H. M. Nussenzveig [9] which aptly summarizes much of the subject matter addressed in this article.

> The rainbow is a bridge between two cultures: poets and scientists alike have long been challenged to describe it. The scientific description is often supposed to be a simple problem in geometrical optics.... This is not so; a satisfactory quantitative theory of the rainbow has been developed only in the past few years. Moreover, that theory involves much more than geometrical optics; it draws on all we know of the nature of light....
>
> Some of the most powerful tools of mathematical physics were devised explicitly to deal with the problem of the rainbow and with closely related problems. Indeed, the rainbow has served as a touchstone for testing theories of optics. With the more successful of those theories it is now possible to describe the rainbow mathematically, that is, to predict the distribution of light in the sky. The same methods can also be applied to related phenomena, such as the bright ring of color called the glory, and even to other kinds of rainbows, such as atomic and nuclear ones.

Acknowledgements

their book. I thank Raimondo Anni for his helpful comments at the proof stage of this article.

Note: For information about recent developments in atomic, molecular, and nuclear scattering, visit `http://cl.fisica.unile.it/~anni/Rainbow/Rainbow.htm`.

References

[1] J. A. ADAM, The mathematical physics of rainbows and glories, *Phys. Rep.* **356** (2002), 229–365.
[2] M. V. BERRY and C. UPSTILL, Catastrophe optics: morphologies of caustics and their diffraction patterns, *Progress in Optics*, vol. 18 (E. Wolf, ed.), North-Holland, Amsterdam, 1980, pp. 257–346.
[3] C. CHESTER, B. FRIEDMAN, and F. URSELL, An extension of the method of steepest descents, *Proc. Cambridge Philos. Soc.* **53** (1957), 599–611.
[4] W. T. GRANDY, *Scattering of Waves from Large Spheres*, Cambridge University Press, Cambridge, 2000.
[5] H. KRAGH, Ludvig Lorenz and nineteenth century optical theory: The work of a great Danish scientist, *Appl. Optics* **30** (1991), 4688–95.
[6] N. A. LOGAN, Survey of some early studies of the scattering of plane waves by a sphere, *Proc. IEEE* **53** (1965), 773–85.
[7] R. G. NEWTON, *Scattering Theory of Waves and Particles*, Springer-Verlag, New York, 1982.
[8] H. M. NUSSENZVEIG, High-frequency scattering by a transparent sphere. I. Direct reflection and transmission, *J. Math. Phys.* **10** (1969), 82–124; High-frequency scattering by a transparent sphere. II. Theory of the rainbow and the glory, *J. Math. Phys.* **10** (1969), 125–76.
[9] ——, The theory of the rainbow, *Sci. Amer.* **236** (1977), 116–127.
[10] ——, *Diffraction Effects in Semiclassical Scattering*, Cambridge University Press, Cambridge, 1992.
[11] D. B. PEARSON, *Quantum Scattering and Spectral Theory*, Academic Press, London, 1988.

About the Cover

Supernumerary Bows along with Airy's Explanation

This month's cover accompanies John Adam's review of *The Rainbow Bridge*. Both images are taken from the book (figures 8-3 and 8-11). The top photograph was taken in 1979 near Kootenay Lake, British Columbia, by Alistair Fraser, one of the book's authors. It is one of hundreds of rainbow photographs he has taken in his life, and shows well a set of supernumerary bows. At the bottom is a plot of rainbow color configurations versus rain drop size, with the size of the drops in the particular rainbow at the top indicated by the slice. Smaller drop radii are at the left, and the slice corresponds to a radius of about a quarter of a millimetre.

The significance, perhaps even the reality, of these 'extra' bows was not appreciated for a long time, quite likely since they were not something Newton's theory could account for. Even today, few people seem to be aware of them. They are faint, but not difficult to observe if one looks carefully.

Alistair Fraser tells us that the historical role played by such bows in motivating the development of the wave theory of light is greater than usually suggested in physics books. He also points out that in nature rain drops are rarely homogeneous in size or shape, and that this has the effect that supernumerary bows are seen only in the top of the bow, where interference arising from the variety of drops does not cancel. He says further, "As supernumerary bows are an interference pattern strongly dependent upon drop size, and as rain showers have a wide range of drop sizes, one would not expect these bows to be seen in showers. They are, nonetheless, seen. Curiously, not all drops contribute equally—through a mixture of the physical optics of small drops and the non-spherical shape of large drops, a narrow band-pass filter is produced which eliminates the supernumerary bows produced by all but a small range of sizes."

The color map is that produced by the theory of George Biddell Airy, dating from roughly 1825. The book includes also a similar map (figure 10-29) deduced from the more exact Mie theory, but emphasizes that the fine detail this predicts, although confirmed under laboratory conditions, is imperceptible in nature, where perturbations blur it.

—*Bill Casselman*
(covers@ams.org)