Mathematical Explanations of the Rainbow

Christopher Pincock
pincock@purdue.edu

May 11, 2010
9590 words
DRAFT

Abstract

Explanations of three different aspects of the rainbow are considered. The highly mathematical character of these explanations poses some interpretative questions concerning what the success of these explanations tells us about rainbows. I develop a proposal according to which mathematical explanations can highlight what is relevant about a given phenomenon while also indicating what is irrelevant to that phenomenon. This proposal is related to the extensive work by Batterman on asymptotic explanation with special reference to Batterman's own discussion of the rainbow.

I. For some time Batterman has used examples from contemporary physics to argue that asymptotic reasoning poses a challenge to traditional accounts of explanation, reduction and emergence. The core instance of this style of reasoning is the transformation of one representation into another via the operation of taking a limit of some quantity, often to 0 or infinity. A central theme of these arguments is that philosophers have restricted themselves to a spartan logical framework which is ill-equipped to do justice to the subtle interplay between the physical and mathematical concepts deployed in our best scientific representations. Batterman has also criticized the attempt to move beyond a purely logical framework by privileging causal criteria. This is because asymptotic reasoning cannot be understood unless we take account of the non-causal aspects of our representations which make contributions to scientists' judgments about these matters.

While I am very much in agreement with Batterman on these points, I worry that he has not sufficiently clarified the links between the many different features of asymptotic reasoning which he has discussed and his positive conclusions concerning explanation, reduction and emergence. Part of the difficulty is that many of his examples involve scientific theories whose interpretative status remains controversial, to say the least. For example, the role of "semi-classical" reasoning in quantum mechanics is hard to get a handle on precisely because there is no agreement on what to say about "ordinary" cases of quantum mechanics or how they relate to traditional reasoning in classical mechanics. In a similar fashion, although surely to a lesser extent, the representation of phase transitions using statistical mechanics and concepts from classical thermodynamics involves some interpretative controversy, most notably in connection with probability. All this suggests that Batterman will have a tough time isolating what is distinctive about asymptotic reasoning if he appeals only to these sorts of cases. What he presents as a case of emergence, for example, may be accounted for in other terms based on a competing interpretation of what is going on.

I will try to clarify the role of asymptotic explanation using a case which Batterman has discussed which is relatively free of interpretative controversy. This is the rainbow. As Batterman has put it, "the problem of explaining the basic features of rainbows is less encumbered with philosophical puzzles than is the interpretation of quantum mechanics" (Batterman 1997, p. 395). This is due, at least in part, to our comfort with the most fundamental theory involved in this explanation. This is classical electrodynamics and the associated conception of light as an electromagnetic wave. But Batterman has argued that our best representation of the rainbow also deploys the ray representation which treats light in terms of geometric lines. The ray representation can be seen to result from the wave representation through the operation of taking a limit. On a first pass, we can say that the rays of the ray theory result from the waves of the wave theory when the wavelength is taken to 0.[1] Following Batterman, we will see to what extent the ray representation is able to contribute to our best explanation of certain features of the rainbow. This is initially puzzling as it is hard to see how the clearly false claim that light travels along rays can help to explain aspects of the rainbow. A further important part of the rainbow case is that some features of the

[1]We will return to the question of whether or not this is the best way to think about the relationship between these two representations.

rainbow are explained by combining elements from the ray representation and the wave representation. This involves what can be called "intermediate asymptotics" and is relevant to our best explanation of the existence and spacing of the part of the rainbow known as supernumerary bows. An appeal must be made to what happens "on the way" to the limit even though the limit operation does not correspond to any physical process or aspect of the rainbow as it really is. The mathematics of the representation allows considerable explanatory insight based on an appeal to "structures" whose special status seems to depend on both the ray and wave representations.

Batterman has summarized the situation by saying that "It seems reasonable to consider these asymptotically emergent structures to constitute the ontology of an explanatory "theory", the characterization of which depends essentially on asymptotic analysis and the interpretation of the results. This full characterization must make essential reference to features of both the ray and wave theories" (Batterman 2002, p. 96). But both sympathizers and critics have not agreed on what Batterman takes the interpretative significance of this "theory" to be. To help to clarify the situation I will make use of a distinction between representation and theory which Batterman does not employ. On this approach, the theories we accept are those which tell us what a system is made up and how these parts interact. I will argue that even though we use representations which were originally derived from the ray theory, their role in our best explanations need not lead to an acceptance of the ray theory. Instead, the only theory we accept is the wave theory. But this is consistent with an important role for the ray representation and its interpretation in shaping our beliefs about the rainbow. I believe that this conclusion is completely in the spirit of Batterman's many remarks on the interpretive implications of asymptotic reasoning. So, I do not intend my discussion in this paper as a criticism of his views. At the most, what I am arguing is that Batterman's views are not as clear as they should be, and that this has hampered the appreciation of the significance of cases like the rainbow.

An important conclusion of this paper is that it is helpful to distinguish two different explanatory benefits associated with asymptotic reasoning. The first benefit is that such reasoning allows us to ignore details which are not relevant to the phenomenon being explained. This sort of contribution obtains for many ways of deploying mathematics in scientific reasoning and has nothing specific to do with the taking of limits.[2] The second benefit can be more directly tied to asymptotic reasoning and can also be used to help motivate Batterman's emphasis on the existence of the singularities which sometimes arise when limits are taken. This benefit is that the representations which result from such limits must be given a different physical interpretation than the original representations. This marks an important difference from a case like wave dispersion where asymptotic reasoning also plays an important role.[3] There the representation which resulted from taking the limit did not require a new interpretation. Now, as we will see, the singular character of the limit can be linked to the need to offer an interpretation in terms of different physical concepts. The challenge then becomes to say how different this physical interpretation must be and what consequences this has for our overall conception of the phenomenon being represented. I will argue that in the rainbow case, at least, some of the strong interpretative conclusions that Batterman has suggested should be weakened. We use the properly interpreted ray representation to help us interpret other representations which arise in the course of asymptotic analysis. As I will put it, we use the results of one idealization to inform the proper interpretation of another idealization. This shows how concepts which were originally deployed in the context of the ray theory are relevant to our best explanations. But it does not show that we need to accept any aspects of the traditional ray theory when we consider the rainbow. Again, I will emphasize that I take my exposition of the details of the case to be primarily a clarification of what Batterman has offered, and not a criticism. But in the concluding section I will argue that Batterman's suggestion that some of our explanations are " 'theory laden' " (Batterman 2005, p. 159) with respect to the ray theory is misleading. Throughout I will also indicate how how a certain approach to idealization is crucial to avoiding interpretative mystery.

II. Rainbows can be initially characterized in terms of the following observable features. They appear when an observer looks towards rain with the Sun behind her. Rainbows involve a relatively bright band centered on a 42° angle above the direction of sunlight. An initially puzzling feature is that this angle remains the same as the observer moves towards or away from the rainbow. The band itself is divided into colors with red, orange and yellow

[2]Cases like this are the focus of my "Mathematical Contributions to Scientific Explanation".

[3]See my "A Recipe for Avoiding Inconsistent Idealizations".

Figure 1: A rainbow.

appearing at the top followed down through the spectrum to blue and violet at the bottom. Finally, in the blue-violet region some rainbows exhibit dark bands. The spacings of light and dark in this region of the rainbow are the supernumerary bows, while the rainbow itself can be characterized as the primary bow.[4] Given these many features of a rainbow and the reoccurrence of the rainbow across various conditions, there are many different explanatory questions that a scientist can ask. I will focus on three: (a) Why does the rainbow appear at 42° relative to the direction of sunlight, when it does appear? (b) Why does the color pattern of the rainbow appear as red on top through violet at the bottom? (c) Why do dark lines appear at the bottom of the rainbow in the blue-violet region with this or that pattern of spacing? These explanations involve different relationships between our fundamental wave theory and the ray theory.

A purely ray theoretic explanation seems available for (a).[5] It involves the idea that the rainbow results from light rays passing through individual raindrops. Crucially, we can account for the 42° angle if we assume that the raindrops are spherical. Based on a mathematical argument to be given below, the spherical shape of the raindrop has the effect of focusing the sunlight

[4]Considerations of space preclude me from discussing the secondary bow which sometimes appears above the primary bow.

[5]My exposition is based mainly on Nahin 2004. See also Adam 2003 and Nussenzveig 1977.

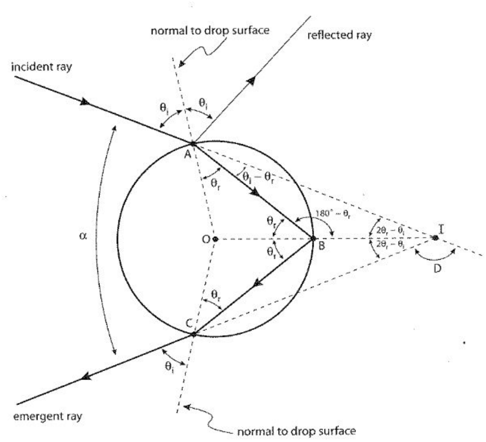

Figure 2: Nahin 2004, p. 59.

so that much of the light that hits each drop is directed backwards and downwards towards the observer. The perceived light at 42° thus corresponds to light that has undergone a minimal total deflection of 138°. This "rainbow angle" θ_R obtains for each spherical raindrop, independently of its size, and so the accumulated effect is the band of light perceived by the observer.

The main physical assumptions underlying this explanation of (a) are Snell's law concerning the refraction of light as it passes from one medium to another and the claim that the angle of reflection equals the angle of incidence when light is reflected. What we want to calculate is how the deflection of a ray of light from the sun changes as we consider rays which strike the drop. As figure 2 shows, a path that involves an initial refraction, followed by a reflection on the back of the drop, followed by a second refraction can lead to a ray that will be deflected backwards and downwards towards the observer. The angles in question here are the initial angle of incidence θ_i and the initial angle of refraction θ_r. Snell's law and the geometry of the circle allow us to determine the total angle of deflection D using just θ_i:

$$D = 180° + 2\theta_i - 4\sin^{-1}(\frac{1}{n}\sin\theta_i) \qquad (1)$$

The basic idea of our explanation of (a) is that (1) tells us that much of the sunlight that hits the raindrop at angles $0° < \theta_i < 90°$ will have D at or close to $D = 138°$. This corresponds to the observed angle of $\alpha = 42° = 180° - D$

that we are trying to account for. This proposal can be made more precise by noting that D obtains a minimum, and so α obtains a maximum, as we move from $\theta_i = 0°$ to $\theta_i = 90°$ when $D = 138°$. Furthermore, this minimum corresponds to a large amount of light because many θ_i lead to this value for D or something close to it. 42° is the maximum value for α, so according to this explanation no sunlight from these rays appears above the primary rainbow.

So far no mention has been made of the colors of the rainbow. That means that our explanation of (a) is clearly not an explanation of (b). But a small addition to our account of where the rainbow appears is sufficient to rectify this gap. In the above argument we treated only the case where $n = \frac{4}{3}$. Treating n as a constant is sufficient to explain where the rainbow appears, but a more accurate perspective on n is that it is a variable which depends not only on the media involved, but also on the color of light undergoing the refraction. At one end of the visible spectrum, the value of n for red light is roughly 1.3318, while for violet light at the other end of the spectrum, n is approximately 1.3435. This slight change in n can be used to account for the angular spread between the colors of the rainbow and their order. For example, our new value for n for red light yields a value of D to be 137.75°, while with the new value of n for violet D becomes 139.42°. As a result, α is greater for red than for violet and the red band appears above the violet band in the sky (Adam 2003, p. 89).

While this is clearly *an* explanation for (b) which can be easily combined with our explanation of (a), it is not clear that this is the *best* explanation of (b). Furthermore, when we press on this explanation for (b) and see an explanation which has a better claim on being the best explanation for (b), we will encounter the worry that our explanation of (a) is problematic as well. To see the limitations of our explanation of (b) notice that we have treated the link between color and n as a series of brute facts. It is here that an appeal to the wave theory of light promises to do better. For once we make the link between the decreasing wavelength of light and the change in color from red through the visible spectrum to violet, we can account for the change in n. According to the wave theory of light what is going on here is the more general phenomena of light dispersion where the behavior of light can be affected by the wavelength of the light wave. A beam of white light is then represented as a superposition of waves of various wavelengths which are separated as each component wave is refracted at a slightly different angle. Going further, we can use this wave representation to account for the physical claims which the ray representation takes for granted. These claims include Snell's Law and the law of reflection.

On this view the best explanation of (b) takes light to be made up of electromagnetic waves. This is the best way to make sense of the colors of light and how the colors come to arrange themselves in the characteristic pattern displayed by the rainbow. Now, though, we have a reason to doubt the explanation we have given for (a) because this explanation presented light as traveling along geometric rays. Without some relationship to our wave representation of light, this account of (a) seems to float free of anything we should take seriously. This point raises an instance of the central question of this paper: how can we combine the resources of the clearly incorrect ray representation with the correct wave representation to provide our best explanations of features of the rainbow? The answer that I would like to suggest follows immediately from a certain approach to idealization. This is that the ray representation we have deployed need not require the assumption that light is a ray. It may assume only that in certain circumstances *some features* of light can be accurately represented using the ray representation. But what are those circumstances and how do they relate to these features? A first attempt to answer this question might draw attention to one of the mathematical relationships between the two representations. This is that the ray representation results from the wave representation when the wavelength is taken to 0. Another way to put this makes use of the wave number k, which is equal to 2π divided by the wavelength. We then can put this limit in terms of the wave number going to infinity. This leads to a representation in which the wave-like behavior of light is completely suppressed. We can then track the path of the light using geometric lines and deploy simple rules like the law of reflection. The lines result from connecting the crests of the light waves. As the wavelength is decreased, these crests get closer and closer and at the limit produce a line.

The problem with this first attempt is that this sort of relationship between the wave representation and the ray representation would also block any ray-theoretic representation of the connection between color and the index of refraction n. For if we interpret the scope of our ray representation as going along with the $k \rightarrow \infty$ limit, then we have erased the very variations in the light waves which are responsible for color and the changes in n. It is hardly coherent to say that if the wave number is close to infinity, then we can offer our explanations of (a) and (b). For, to start with, the wave number is not close to infinity in any absolute sense. More importantly, the differences between the wave numbers of red and violet light are manifestly part of the explanation for the colors of the rainbow. So we cannot defend the view that it is correct to assume that the wave number is infinity even if this would render the links between the ray and wave representations particularly transparent.

The lesson we should draw from the failure of the first attempt is that we need to pay attention to both the mathematical links between the two representations and their proper physical interpretation. Some mathematical links will preclude any viable interpretation. My proposal here is that the circumstances under which our ray representation can be used to provide explanations of (a) and (b) should be specified in terms of a dimensionless parameter known as the size parameter β. β is the product of the wave number k with the radius a of the raindrop. It is a dimensionless quantity because its value is independent of the units used to represent k and a. On this view, the ray representation results from the wave representation when the product ka is taken to infinity.[6] This ray representation can be reliably used if there is a size of raindrop above which the wave-theoretic aspects of light are not relevant to the path of the light through the drop. If the raindrops have a circumference of roughly 1 mm, then the product in question here is about 5000 (Nussenzveig 1977, p. 124). This makes it possible to treat the paths of light as straight lines and to draw on the link between the wave number and the index of refraction. For even though we have assumed that $\beta \to \infty$, this assumption is consistent with appealing to the values of k which are responsible for the dispersion of the light. Analogously, we do not represent the wave crests as forming a continuous straight line, but only claim that the distance between crests is so small with respect to the radius of the drops that it is not relevant to the path of the wave.

III. The conclusion of the previous section is that we can account for our best explanations of (a) and (b) by linking the wave representation to the ray representation using a limit specified in terms of the size parameter β. There are two ways in which this two-legged approach provides superior explanations to a one-leg approach that tried to get by with just the wave representation by itself. The most obvious advantage is that we see that many detailed wave representations of a specific rainbow will be related to a single ray representation. This is because the limiting relationship drops

[6] When he is being careful Batterman will also invoke this parameter, and not simply the wavelength. See, e.g., Batterman 2005, p. 154 and Batterman 2010, p. 21, fn. 23

many details which are not relevant to the features of the rainbow which we are trying to account for. Our best explanations of (a) and (b) aim to account for features which all rainbows have in common and we come to understand how these features arise independently of many features which would be included in a viable wave representation of a particular rainbow. Here the parameter used to link the wave representation to the ray representation can be seen to have special significance. For the successful explanation which results from taking $\beta \rightarrow \infty$ shows that it is the size of this parameter which is crucial to these features. When this parameter fails to be large, we can then also explain why these features fail to materialize. For example, if the rain drops are too small, then a rainbow will not be observed or its colors will be distorted.

The two-legged explanation gains part of its explanatory power from its function of unifying our description of the different rainbows at the right level of abstraction. This contribution to explanatory power can be tied directly to the mathematical operation which links the wave representation to the ray representation. But it would be a misunderstanding of this sort of contribution to think that a unified description at the right level of abstraction must result, or even usually results, from applying these sorts of limits. Another means to obtain just this sort of explanatory contribution from mathematics is to attempt to describe a phenomenon directly at the right level of abstraction. So, this sort of explanatory contribution from mathematics need not involve limits for the simple reason that we may not articulate two different representations of the sort deployed in the rainbow case.

Still, there is a second sort of mathematical contribution to the explanatory power of our explanations of (a) and (b) which can be more directly tied to the presence of our limit operation. This is the way in which the mathematical character of the limit operation guides the physical interpretation of the ray representation. If we consider the ray representation independently of its relationship to the wave representation, then we may be tempted to assign physical significance to its parts in line with the traditional ray theory of light. But when we view the ray representation as the product of our limit operation, then we see that this traditional interpretation is problematic. To take a feature of the rainbow that will assume greater significance shortly, the relative intensity of the light which makes up the parts of the rainbow is not accurately captured by our ray representation. We see this when we consider what happens to the wave representation's way of handling intensity when the limit is applied. The intensity is calculated by squaring the amplitude of a light wave. But if we hold the energy transmitted by a light wave fixed and decrease the wavelength, then the amplitude will increase. Therefore, if the wavelength goes to zero, the intensity of a given light wave will go to infinity! This problem is not really avoided by the change to the size parameter β. For we still have no correlate in the ray representation to the intensity of a light wave in the wave representation. This shows how taking the limit operation can have positive as well as negative effects. The positive effect is that we erase the representation of irrelevant details, while a negative effect is that we erase the means to represent the intensity of a ray of light using the link to the wave representation.

So, limit operations force us to withdraw the physical interpretation from some parts of the representation which results. In this respect they are like many other techniques of idealization and simplification. The need to withdraw an interpretation should be obvious for the parameter β. If we take it to infinity to derive our ray representation, then we cannot make use of its actual finite values when we interpret the ray representation. But sometimes other aspects of the representation will be affected in a more subtle way so that some care is needed to ascertain which parts of the resulting representation retain their original significance. This is especially important to keep in mind when the limiting operation produces singularities, i.e. quantities which diverge or become otherwise undefined. Intensity is just such a quantity. If we over-interpret the ray representation, then our explanation of (a) and (b) also involves the claim that the intensity of the rainbow is infinite. But if we judiciously refrain from interpreting this aspect of the ray representation, then we can provide our explanation and retain coherent commitments about the system being represented.

How, though, does this sort of interpretative shift contribute to explanatory power? My suggestion is that the interpretive flexibility that is necessary to handle these shifts helps us to explain *other* aspects of the rainbow which would otherwise remain inexplicable. This comes out clearly in our best explanation of (c) the existence and spacing of the supernumerary bows. As I will summarize in the next section, the best explanation of (c) is obtained by combining the wave representation and the ray representation in a more nuanced way than what was required for our explanation of (a) and (b). Roughly, the best explanation of (c) will result from using information from the ray representation to guide the interpretation of a new third representation. This guidance seems to draw on the novel interpretations which we assign to the ray representation over and above what is grounded in the wave

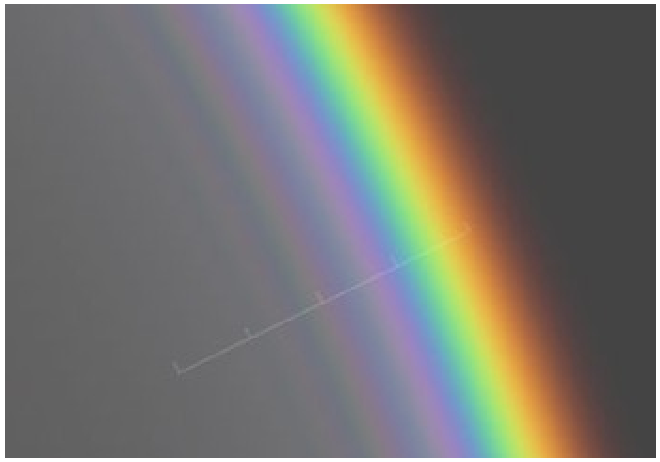

Figure 3: Rainbow with supernumerary bows.
http://www.atoptics.co.uk/rainbows/

representation. So, it seems that the ray representation is making a genuine contribution to the explanation. The central issue, though, is what interpretative significance this has. After considering the explanation of (c) we will turn to Batterman's and Belot's contrasting views on its significance for our beliefs about the rainbow.

IV. The best explanation of (c) the existence and spacing of the supernumerary bows involves both the wave theory and the ray theory in a more intricate way than what we saw in the case of (a) and (b).[7] The supernumerary bows result from the constructive and destructive interference of the light waves which travel through a raindrop. While this satisfactory qualitative explanation was arrived at in the nineteenth century, as late as 1957 the scientist van de Hulst could complain that a "quantitative theory of the rainbow for the broad range of size parameters occurring in nature was still lacking" (Nussenzveig 1992, p. 29).[8] This was despite the fact that a representation was formulated in 1908 by Mie which described the situation adequately in terms of electromagnetic theory. The problem with the Mie representation is that it involves an infinite series of terms which converges very slowly. The development of computers allowed a reliable estimate of the values of this

[7]My main sources are Nussenzveig 1992 and Adam 2002.
[8]See van de Hulst 1957, p. 249.

sum for different parameters, but it was widely felt that this sort of brute force calculation was inconsistent with genuine scientific understanding. As Nussenzveig puts it, "a computer can only calculate numerical solutions: it offers no insight into the physics of the rainbow" (Nussenzveig 1977, p. 125).[9] An adequate explanation was achieved only when Nussenzveig and his collaborators applied sophisticated mathematical techniques which transformed the slowly converging series of the Mie representation into a more transparent representation. The resulting "complex angular momentum" approach or CAM provides the best explanation of the existence and spacing of the supernumerary bows.[10] My purpose here, then, is to consider what interpretative significance this explanation has. In particular, given the acceptance of some form of inference to the best explanation, what should we infer about rainbows given this explanation? My ultimate conclusion is that the mathematical techniques of CAM are central to the superior explanatory power of the theory because the techniques depend on a particular physical interpretation of the mathematics. CAM is thus best seen as radically different from a more straightforward numerical calculation. The "insight into the physics of the rainbow" which results has implications for what we should believe about the rainbow. At the same time, it does not provide us with a distinct theory which should shape our beliefs about the constituents and interactions which give rise to the phenomenon. The only genuine theory which we are left with is the electromagnetic theory.

The explanation offered by CAM is obtained from the Mie representation in three steps. The Mie representation uses electromagnetic theory to describe what happens when monochromatic light, represented as a plane electromagnetic wave, is scattered by a spherical, homogeneous drop of water.[11] The central parameter of the problem is the size parameter β encountered earlier. A further familiar parameter is n, the index of refraction. For the air to water case the value typically employed is 4/3. The problematic part of the Mie approach is the representation of the *scattering amplitudes* at a given angle θ. The amplitude is given by two terms $S_j(\beta, \theta)$ $(j = 1, 2)$ corresponding to two perpendicular directions of polarization. Each $S_j(\beta, \theta)$ term is expressed as an infinite sum of terms for $l = 1$ to $l = \infty$ which involve what are called partial waves $S_l(\beta)$ as well as p_l terms which are defined using the

[9] Quoted at Batterman 1997, p. 407.

[10] Although CAM is sometimes referred to as "CAM theory", I will drop the word "theory" to avoid confusing CAM with the restricted sense of "theory" defined earlier.

[11] A steady-state approach is deployed from the start here.

Legendre function. The partial waves $S_l(\beta)$ are identified with a complicated fraction involving Bessel and Hankel functions (Nussenzveig 1992, p. 39).[12] A reliable numerical calculation of the S_j functions requires an evaluation of around $\beta + 4\beta^{\frac{1}{3}} + 2$ terms (Nussenzveig 1992, p. 40). For an ordinary rainbow, β may exceed 4500, so more than 4500 terms may need to be considered before the calculation is terminated. Accurate predications concerning the supernumerary bows result, but the explanatory power of this calculation is limited.

The first step away from the Mie representation is to transform the infinite sums for each $S_j(\beta, \theta)$ into what is known as a Debye expansion. This step can be motivated by considering the account of the rainbow offered by the ray representation. The light that we see results from an initial refraction, followed by an internal reflection and then by a second refraction. But some light from the initial refraction will be reflected and some light at the internal reflection point will be refracted out of the rain drop. We can track these interactions by considering the reflection coefficients R and the transmission coefficients T. This allows us to rewrite a partial wave term $S_l(\beta)$ as a sum of terms with index $p = 0$ to $p = \infty$. p here indicates how many internal reflections are being considered:

$$S_l(\beta) = S_{l,0}(\beta) + S_{l,1}(\beta) + S_{l,2}(\beta) + \Delta S_{l,P}(\beta) \tag{2}$$

$S_{l,0}(\beta)$ tracks how much light from this term will be directly reflected, while $S_{l,1}(\beta)$ captures the light that is directly transmitted through the drop. Crucially, $S_{l,2}(\beta)$ tracks the light that is reflected internally once. By appeal to the ray representation, we know that this is the light that makes up the rainbow. The remaining terms are grouped together as $\Delta S_{l,P}(\beta)$. They are not relevant for the rainbow, but could be examined in more detail for other phenomena where they become important.

Substituting the new representation of the partial waves into the sum provided by the Mie representation yields the Debye expansion of the total scattering amplitudes (Nussenzveig 1992, p. 92):

$$S_j(\beta, \theta) = S_{j,0}(\beta, \theta) + S_{j,1}(\beta, \theta) + S_{j,2}(\beta, \theta) + \Delta S_{j,P}(\beta, \theta) \tag{3}$$

The values for the transmission and reflection coefficients for water and air

[12] These are some of the so called “special functions”. See Batterman 2007 for some discussion.

show that nearly all of the light which hits the drop is found in the first three terms of this Debye expansion.[13]

This is some progress, but much of the complexity of the original Mie representation is now simply packed into the evaluation of the first three terms of the Debye expansion. The way forward is to apply our second step, namely to make an appeal to the Poisson sum formula. This formula allows us to identify a sum of infinitely many terms with a sum of terms involving an integral (Nussenzveig 1992, p. 45):

$$\sum_{l=0}^{\infty} \phi(l + \frac{1}{2}, x) = \sum_{m=-\infty}^{\infty} (-)^m \int_0^{\infty} \phi(\lambda, x) \exp(2im\pi\lambda) d\lambda \qquad (4)$$

Crucially, we require that the ϕ on the right-hand side be a complex-valued function which agrees with the real-valued function that we started with on the left-hand side when $\lambda = l + \frac{1}{2}$ $(l = 0, 1, 2, \ldots)$. Applying the Poisson sum formula to the first three terms of our Debye expansion might seem to make the problem even worse as we are introducing a complex-valued function and are now required to evaluate an integral involving that function. It turns out, though, that it can be much easier to evaluate an integral involving this sort of complex-valued function. The main reason for this is that there is considerable freedom in changing the path along which we integrate our function. This freedom allows us to choose a path where the dominant contributions to the integral become manifest.

The use of the Poisson sum formula and the consequent use of complex analysis to evaluate the terms of the Debye expansion makes clear what the word "complex" refers to in the complex angular momentum approach. But the interpretation of λ as some kind of angular momentum remains opaque and until it is clarified the whole procedure risks collapsing into some form of numerical calculation. The link is made by a claim known as the localization principle. It associates each value of l in the Mie representation sum with a ray with the impact parameter $b_l = (l + \frac{1}{2})/k$.[14] Each impact parameter is the distance from an axis that passes through the center of the drop (Figure 4). So, when $b_l < a$ the light will hit the drop. By analogy with similar "scattering" problems in particle physics, we can treat the way in which the impact parameter leads to a given deflection angle θ on the model of a

[13] Nussenzveig gives the amount as 98.5 percent (Nussenzveig 1992, p. 95).

[14] Nussenzveig 1992, p. 8. The principle's use in optics is summarized at van de Hulst 1957, pp. 208-209.

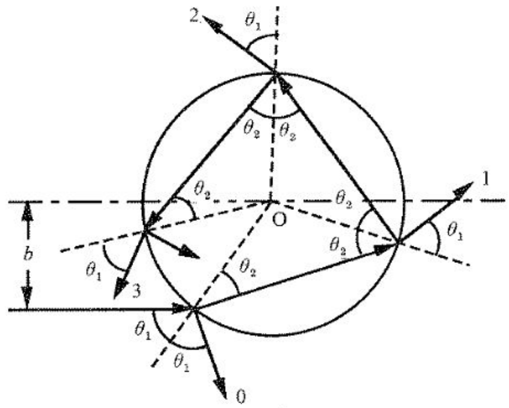

Figure 4: Nussenzveig 1992, p. 16.

particle like an electron being shot through a spherical force field. This makes it useful to think of λ in terms of the angular momenta of the components of the light which hits the drop.

The third step of our explanation involves evaluating the three terms of the Debye expansion which have been transformed by the Poisson sum. For ease of exposition I consider only the third term $S_{j,2}(\beta, \theta)$ corresponding to one internal reflection, but the same techniques can be used to evaluate the other terms as well. Here we have an integral involving the complex function $S(\lambda, \beta)$ which extends the real function $S(l, \beta)$. The basic idea of evaluating these integrals is that we need only consider the contributions from critical points. These critical points come in two kinds. First, there are *saddle points* where the first derivative of S with respect to λ is 0.[15] Second, there are *poles* where S lacks a derivative of some order. In the simplest kind of case, as with the function $\frac{1}{z}$ at the point $z = 0$, the function is undefined. To evaluate an integral with a complex function it is often sufficient to know the saddle points and the poles. The trick is to consider a path of integration which departs from the real axis, but which goes through the saddle points, as in figure 5.[16] Drawing on these facts, we can then see how the value of the integral changes as we vary θ by considering how the saddle points

[15]This is a saddle point, and not a minimum or maximum, because the second derivative changes sign around this point.

[16]Saddle points are indicated by circles, while poles correspond to crosses.

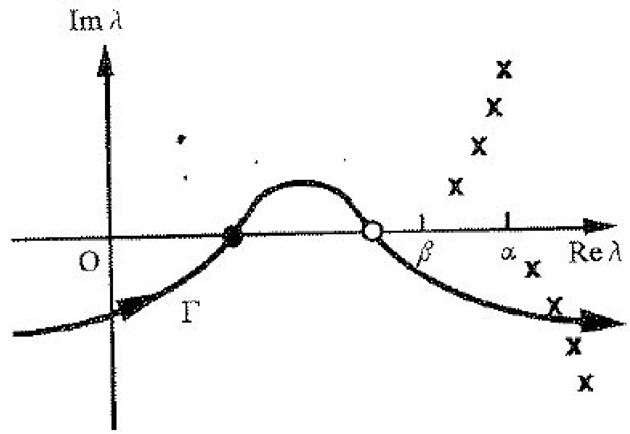

Figure 5: Nussenzveig 1992, p. 104.

change. Unsurprisingly, an important change occurs when θ is close to θ_R, the rainbow angle. When θ is π radians or 180°, there is a single saddle point at $\lambda = 0$. As θ decreases, the saddle point moves to the right along the real axis. There is then a critical value θ_L at which a second saddle point appears even further down the real axis at $\lambda = \beta$. Finally as θ gets closer to θ_R, these two saddle points approach closer to each other. At $\theta = \theta_R$ the two saddle points merge.[17] If θ is decreased further, then the saddle points diverge from each other and move into the complex plane.

The physical interpretation of these saddle points is crucial to an evaluation of the CAM explanation of the supernumerary bows. The basic idea behind this interpretation is that the saddle points correspond to geometric rays. The existence of a single saddle point in the $\pi > \theta > \theta_L$ region indicates that only a single geometric ray appears at these deflection angles. But when $\theta_L > \theta > \theta_R$ the two saddle points correspond to two geometric rays combining to increase the amount of light scattered in this direction. This focusing effect is most dramatic when $\theta = \theta_R$ when the saddle points coincide. As Nussenzveig puts it, "a rainbow corresponds to the collision between two saddle points" (Nussenzveig 1992, p. 102). Finally, for $\theta_R > \theta$ the absence of real saddle points shows that there are no corresponding geometric rays.

[17]Here I omit some complications which this merging causes. See Nussenzveig 1992, p. 106.

The justification of this interpretation of the saddle points is that we are assuming a situation where the dominant contributions to the light arise from electromagnetic waves which approach the behavior of rays of light. In particular, we have assumed throughout that we are operating in a context where the wavelength of light is much smaller than the radius of the drop, the only other relevant length parameter. So, in terms of our size parameter, $\beta \gg 1$. Our mathematical theory tells us that the dominant contributions to our integral will come from saddle points. So, as with the localization principle, there is an interpretative conjecture that these points correspond to rays which would appear more and more sharply if the ratio between the drop radius and the wavelength increased further. This conjecture is supported by some preliminary mathematical analysis of the Mie representation where we find that the angle of deflection of the dominant term corresponds to the angle of deflection predicted by the ray theory (Nussenzveig 1992, pp. 9-11). At the same time, it is important to emphasize the experimental nature of this association. It is not dictated by the Mie representation as the Mie representation does not include light rays in its scope. Similarly, the link between rays and these saddle points is not based on anything that the ray theory of light could tell us. This is mainly because we reject the ray theory. But it is also clear that the ray theory has nothing to say about the saddle points which appear in the CAM treatment.

A similar point can be made about the interpretation of the poles of $S(\lambda, \beta)$ which are also used in the evaluation of our integral. The contributions which the poles make, known as residues, correspond to light which has traveled along the surface of the drop for some distance. Again, the link between the features of S and their physical interpretation is initially conjectural. On the side of the physics we recognize that the ray theory cannot be capturing all of the light relevant to the rainbow. In particular, the way in which light is diffracted by a sphere is ignored. So, we conjecture that some light beyond that corresponding to geometric rays is visible in the rainbow and that this light may result from surface waves which are captured by the surface of the drop. On the mathematics side we know that we are able to evaluate our integral using the poles of our function based on the contribution which they make to the amplitude in a given direction θ. This leads us to match the poles with the contributions from surface waves. As with the saddle points, some preliminary mathematical analysis can reinforce this conjecture.

As the Mie representation can be rendered tractable by a computer it

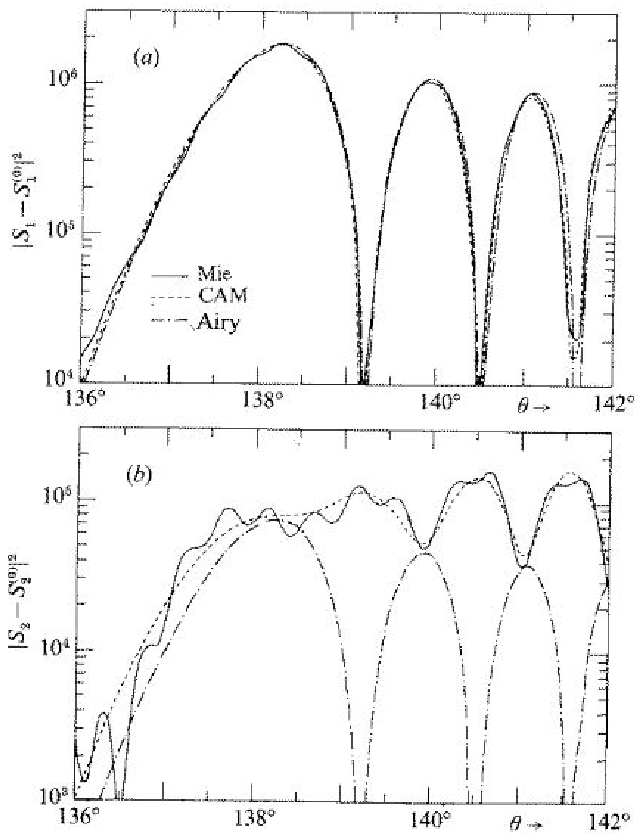

Figure 6: Nussenzveig 1992, p. 113.

is possible to compare the predictions of the Mie representation with the results of the CAM analysis. What we find is that CAM outperforms its predecessors and is able to more or less match the Mie representation for a wide range of the size parameter β. For example, figure 6 shows how CAM compares to the Mie representation and the earlier Airy representation when $\beta = 1500$. These lines represent intensities, which are obtained by squaring the amplitudes. The Airy representation does much better for one direction of polarization than for the other. The dotted line of CAM tracks the Mie representation quite closely, both for the rainbow around 138° and the two additional peaks corresponding to two supernumerary bows. By contrast, the Airy representation incorrectly puts peaks for supernumeraries where there are in fact dips in the light intensity. From a standpoint of predictive accuracy, then, CAM is comparable to the Mie representation.

V. Predictive success is not sufficient for explanatory power, so it is worth asking why CAM is deemed such a good explanation of the existence and spacing of the supernumerary bows. Two factors seem crucial. The first is the way in which each step in the derivation is susceptible to a physical interpretation. As a result, it is possible to adopt the physical interpretation and come to understand why the rainbow has the features it does. For example, the reasons and details of how the size parameter β affects the supernumeraries becomes much clearer. Using the terms of the Debye expansion, we can see that light which is reflected off the drop and light which is internally reflected is relevant to the rainbow. Furthermore, the localization principle and the association between saddle points and rays lets us appreciate how the light behaves in some respects as the ray theory would predict. But the additional role of the poles and their link to surface waves points to the importance of diffraction and the wave character of light even when the size parameter β is large.

A second factor which clearly contributes to the explanatory power of CAM is its wide range of applicability. This range has a number of aspects. First, restricting our focus to a particular rainbow with some fixed β, CAM delivers a "uniform" representation which is valid for the whole range of θ. By contrast, there are "transitional" representations which hold only for limited ranges of θ. While better than nothing, such transitional representations often result from more or less ad hoc adjustments which do not indicate where the representation will fail or how it can be extended. By contrast, a uniform representation will account for the success of more limited transitional representations and perhaps even correct their mistaken interpretations.[18] A related aspect of the range of applicability concerns the acceptable variation of the parameter β. The Airy representation can be used only when β is very large, e.g. 5000. CAM can account for this limitation and is valid in a much

[18]For example, the earlier flawed Fock theory relied on "transverse diffusion" (Nussenzveig 1992, p. 60) as opposed to surface waves. As Nussenzveig emphasizes, the success of CAM over the Fock theory vindicates van de Hulst's earlier proposal that surface waves are important in diffraction, and undermines Fock's proposal (Nussenzveig 1992, pp. 34, 85-86). See van de Hulst 1957, §17.52, drawing on work by Beckmann and Franz noted at p. 227.

larger range, i.e. $\beta > 50$.[19]

There is an additional aspect of the range of applicability which fits with the popular link between explanation and the unification of otherwise disparate phenomena. Other meteorological phenomena, such as the secondary rainbow and the glory, can be successfully treated using CAM. The relationship between the rainbow and the glory, in particular, was a subject of much investigation and it is only with CAM that an adequate explanation of the glory was found.

Finally, there are aspects of CAM which allow it to be extended to other sorts of physical phenomena. This is possible because the mathematics can be interpreted in terms of a completely new range of physical properties. There are many cases of "scattering" such as in particle physics or even acoustics where phenomena which are analogous to the rainbow appear. Thus it is possible to speak of nuclear rainbows and glories and to treat them using the basic framework provided by CAM. This is a different sort of unification of phenomena than the way in which CAM unifies the meteorological rainbow and glory and philosophers seem reluctant to assign it equal importance. Nevertheless, it is an important part of scientific research to search for these commonalities and the insight they provide.

We are now in a position to ascertain what the best explanations of (a), (b) and (c) mandate as far our beliefs about rainbows. My suggestion is that the explanations are successful not just because some mathematical transformations are successfully applied to the Mie representation. As we have seen, the explanations involve a series of conjectures about how the representations which result from these transformations should be interpreted. To the extent that these explanations are the best ones available, it seems reasonable to take these interpretations seriously. This is just a form of inference to the best explanation (IBE). However, in its usual role in the philosophy of science the IBE licenses beliefs in a new kind of entity like electrons. This is clearly the wrong way to think of how IBE is shaping our beliefs in the rainbow case. For the explanations do not posit a new class of entities. Instead, the conjectures involve linking this or that part of the mathematics with features of the rainbow which are deemed to be important for the aspect of the phenomenon being explained. This happened in two different ways. In the explanation of (a) and (b) we considered what occurred if a certain limit was taken. Once this limit was conceived as involving the size parameter β it was possible to

[19] Nussenzveig 1992, p. 112.

see that the explanation did not turn on the assumption that light is a ray as the traditional ray theory would have it. So, this aspect of the traditional ray theory is completely idle in the explanation. At the same time, we came to see how (a) and (b) were due to those aspects of light which could be correctly described by the ray theory. This explanation adds to whatever descriptions of the rainbow are provided by the fundamental electromagnetic theory. This is the sort of "physical insight" which these limit techniques can provide.

Things worked out somewhat differently in the explanation of (c). There we do not consider what happens if β is taken to infinity, but only what happens when β gets sufficiently large. How large is large enough is part of the explanation, and as we have seen CAM proves empirically successful when β gets larger than 50. CAM explains which features of the rainbow play an important role in this context. But now we draw on aspects of both the wave theory and the ray theory. It is not possible to explain the existence and spacing of the supernumeraries using just concepts deployed in the ray theory. This is because interference and diffraction are central to what is being explained. At the same time, it is not possible to explain this phenomenon by appeal only to what we find in the wave theory. Important links were made between the critical points and aspects of the rainbow. These links were given in terms of rays and surface waves. So, they involved aspects of the light scattering which are not available from the perspective of the wave theory alone. Instead, a scientist must ascend from the wave theory to the ray representation before she is able to get the "physical insight" into the supernumeraries which CAM provides. This does not mean that she must believe that the ray theory is correct. Instead, she must use the results of one idealization to inform the proper interpretation of another idealization. The techniques deployed in the explanation of (a) and (b) and their proper interpretation let us explain (c).

VI. In assigning explanatory power to asymptotic techniques and emphasizing the importance of interpretations derived from non-fundamental theories I am of course echoing Batterman's many discussions of these issues. In his first discussion of the rainbow in Batterman 1997, Batterman focuses on Stokes' analysis of the Airy representation. This theory provided at least a partial explanation of (c) for the case when β is very large. But it was Stokes' asymptotic analysis of the Airy function at the heart of the Airy representation which allowed this explanation to succeed. Stokes' techniques are analogous in many respects to the way in which CAM was applied to the later Mie representation. The improved understanding provided by Stokes' analysis leads Batterman to conclude that "these 'methods' of approximation are in effect more than just an instrument for solving an equation. Rather, they, themselves, make explicit relevant structural features of the physical situation which are so deeply encoded in the exact equation as to be hidden from view" (Batterman 1997, p. 396). This is certainly one aspect of CAM as well. So, in this respect we are in agreement with Batterman that these mathematical techniques can contribute explanatory power.

The Stokes case involves the Airy representation and we saw that it has been superseded by CAM. However, in chapter 6 of Batterman 2002, Batterman provides an additional discussion of the rainbow focused on the explanation of (c). His discussion considers the way in which the mathematical theory known as catastrophe theory can be used to obtain many of the results that we have seen follow from CAM.[20] While the mathematical techniques are quite different in appearance, they both involve an analysis of the critical points of functions which result from the Mie representation when a parameter gets large. Crucially, the catastrophe theory analysis concerns how the singularities which emerge in the ray theory behave as the relevant parameters are changed. So, the proper interpretation of the mathematics here involves the ray theory just as much as the CAM techniques do. As Batterman puts it, "one cannot interpret these purely mathematical results in wave theoretic terms alone. An understanding of what the mathematical asymptotics is telling us requires reference to structures that make sense only in the ray theory" (Batterman 2002, p. 96). Unfortunately, Batterman does not say exactly what beliefs about the rainbow should be adjusted based on this analysis. However, everything he does say in his book is consistent with the proposal we arrived at in the last section. This is that the conjectures linking the mathematics to aspects of the ray representation are vindicated by the successful explanation of (c). This gives us "insight" into what features of the light are responsible for (c) and these features are characterized in terms borrowed from the ray representation.

There is a much more deflationary way to read what is going on here and we can see Belot 2005 as taking this route in reaction to Batterman's book.[21] According to Belot,

[20]See Nussenzveig 1992, pp. 115-116 and Adam 2002, §7 for a comparison of the results of the two approaches.

[21]See also Redhead 2004.

> The mathematics of the less fundamental theory is definable in terms of that of the more fundamental theory; so the requisite mathematical results can be proved by someone whose repertoire of interpreted physical theories includes only the latter; and it is far from obvious that the physical interpretation of such results requires that the mathematics of the less fundamental theory be given a physical interpretation (Belot 2005, p. 151).

This may just mean that the success of the mathematical techniques deployed by CAM or catastrophe theory need not lead us to accept any theory besides the fundamental theory. If this is all that Belot intends to say, then he is correct. The insight associated with our explanation of (c) does not involve belief in any *theory* beyond the wave theory. But saying only this underestimates what scientists themselves say about the importance of our best explanation of (c). As we have seen, it shows us which contributions are most important to the supernumeraries, and the importance of these contributions is not accessible from the perspective of the wave theory. Belot's attitude, then, threatens to undermine the crucial distinction between a brute force numerical computation and a well-motivated physical analysis of the situation. The former lacks explanatory power and interpretative significance for our understanding of the rainbow, while the latter increases our scientific understanding by isolating what is responsible for this or that aspect of the rainbow. These important factors are isolated using the interpretation of the ray theory representation arrived at in our explanations of (a) and (b). So, while we do not ever come to adopt a new and competing theory of the rainbow, our insight into the rainbow must be expressed in the vocabulary of representations associated with a non-fundamental theory, in this case, the ray theory.

This is essentially the line that Batterman takes in his reply to Belot. But he also claims that Belot's "pure mathematician" needs the less fundamental theory to motivate the steps which are taken to develop the ray representation in a stronger way than I have suggested. These developments go beyond simply taking a parameter like the wavelength to 0. In particular, Batterman suggests that the initial and boundary conditions of the ray representation make sense only if we interpret them using the ray theory of light:

> those initial and boundary conditions are not devoid of physical content. They are 'theory laden'. And, the theory required to characterize them as appropriate for the rainbow problem in the

> first place is the theory of geometrical optics. The so-called 'pure' mathematical theory of partial differential equations is not only motivated by physical interpretation, but even more, one cannot begin to suggest the appropriate boundary conditions in a given problem without appeal to a physical interpretation. In this case, and in others, such suggestions come from an idealized limiting older (or emeritus) theory (Batterman 2005, p. 159).

If using a theory in this way involves beliefs about the subject-matter, then it seems to me that Batterman has gone too far. It is true that the ray theory originally motivated the ray representation which played a role in the explanation of (a) and (b). Furthermore, I have accepted that this very representation plays a role in the interpretation of the steps leading to our best explanation of (c). But we have not seen a need to adopt any beliefs about light which the traditional ray theory incorrectly offered. For example, the ray theory maintains that light travels along rays. All we have come to accept as a result of our explanations is that in certain contexts there are aspects of light which are accurately captured by the ray representation. The explanatory power of our ray representation turned on the way in which the parts of it which we take seriously can be grounded in the fundamental wave theory. So, the ray theory as it was traditionally presented obscures what is really going on and this is why we reject it. This is consistent with using the ray representation as an idealized representation which conveys accurate information about this or that aspect of the rainbow. And, as we have seen, it is consistent with the best explanation of (c) drawing on certain features of this idealization to guide a further idealization. I believe that this is consistent with Batterman's considered views on the issue, as he says things like "asymptotic explanation essentially involves reference to idealized structures such as rays and families of rays, but does not require that we take such structures to exist" (Batterman 2005, p. 162). But, again, I hope to have presented a clearer view of what these explanations tell us about rainbows.

Perhaps what has led Batterman into these stronger statements is his focus on theories and the traditional assumption that if we deploy a representation that derives from a theory, then this theory must be playing a role in shaping our beliefs about the systems being represented. So, the central role of the ray representation and its interpretation indicate for Batterman that the ray theory must be playing a role in our beliefs about the rainbow.

Belot can also be seen to be making this assumption, but he starts from the fact that we reject the ray theory. So, for Belot, the rejection of the ray theory entails that we cannot really employ any ray representation in shaping our beliefs about the rainbow. If, as I have argued, we reject the traditional assumption linking theory and representation, then we can make sense of the widespread scientific practice of using representations which are associated with outdated theories. These representations help to reveal what is going on in a given situation and are crucial to the explanations which we have. But these explanations are consistent with accepting only the fundamental theories, even though we use concepts from other representations to help us to decide what to believe. Mathematical transformations link these theories to our representations in unanticipated ways and it is these transformations which philosophers must focus on if they are to come to terms with the commitments of practicing scientists.

References

Adam, J. A. 2002, 'The mathematical physics of rainbows and glories', *Physical Reports* **356**, 229–365.

Adam, J. A. 2003, *Mathematics in Nature: Modeling Patterns in the Natural World*, Prentice-Hall.

Batterman, R. 1997, '"Into a mist': Asymptotic theories on a caustic', *Studies in the History and Philosophy of Modern Physics* **28**, 395–413.

Batterman, R. 2002, *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford University Press.

Batterman, R. 2005, 'Response to Belot's "Whose devil? Which details?"', *Philosophy of Science* **72**, 154–163.

Batterman, R. 2007, 'On the specialness of the special functions (the non-random effusions of the divine mathematician)', *British Journal for the Philosophy of Science* **58**, 263–286.

Batterman, R. 2010, 'On the explanatory role of mathematics in empirical science', *British Journal for the Philosophy of Science* **61**, 1–25.

Belot, G. 2005, 'Whose devil? Which details?', *Philosophy of Science* **72**, 128–153.

Nahin, P. J. 2004, *When Least is Best: How Mathematicians Discovered Many Clever Ways to Make Things as Small (or as Large) as Possible*, Princeton University Press.

Nussenzveig, H. M. 1977, 'The theory of the rainbow', *Scientific American* **236 (4)**, 116–127.

Nussenzveig, H. M. 1992, *Diffraction Effects in Semiclassical Scattering*, Cambridge University Press.

Redhead, M. 2004, 'Discussion note: Asymptotic reasoning', *Studies in the History and Philosophy of Modern Physics* **35B**, 527–530.

van de Hulst, H. C. 1957, *Light Scattering by Small Particles*, Wiley and Sons.