

Abstract

This paper attempts to address the problem of the applicability of mathematics in physics by considering the (narrower) question of what make the so-called special functions of mathematical physics special. It surveys a number of answers to this question and argues that neither simple pragmatic answers, nor purely mathematical classificatory schemes are sufficient. What is required is some connection between the world and the way investigators are forced to represent the world.

On the Specialness of Special Functions (The Nonrandom Effusions of the Divine Mathematician)

Robert W. Batterman*

The University of Western Ontario

Special functions include the gamma function (a generalization of the factorial), elementary transcendental functions such as the exponential, the sine, cosine, as well as the “higher” transcendental functions such as, Bessel functions, the Riemann Zeta function, etc. This list goes on and on and, in fact, continues to expand.¹ Michael Berry [11, p. 12] notes, in a paper entitled “Why are Special Functions Special?,” that the persistence of special functions is both surprising and puzzling.

What are they [special functions], other than just names of mathematical objects that are useful only in situations of contrived simplicity? Why are we so pleased when a complicated calculation “comes out” as a Bessel function, or Laguerre polynomial? What determines which functions are “special”? [11, p. 12]

*I would like to thank the philosophy of physics reading group at the University of Western Ontario for comments. In particular, I thank Rob Corless, Bill Harper, Wayne Myrvold, and Brian Woodcock for useful discussions. Thanks also to Michael Berry, Penelope Maddy, and, for very detailed comments and suggestions, Christopher Pincock.

¹The fact that this list continues to grow is indeed an interesting phenomenon in itself. In an age where numerical calculations are virtually trivial on desktop or even handheld computers, it is remarkable that some sort of analytic classification scheme should be desirable. The conclusions of this paper may shed some light on why.

1 Some Responses to the Question

Berry's ultimate, though tentative conclusion, is essentially pragmatic in nature. He recognizes that various extant classification schemes— attempts to unify classes of different special functions in some “natural” order or hierarchy—are incomplete. They typically leave out large classes of other functions. He says,

One reason for the continuing popularity of special functions could be that they enshrine sets of recognizable and communicable patterns and so constitute a common currency. . . . [P]erhaps special functions provide an economical and shared culture analogous to books: places to keep our knowledge in, so that we can use our heads for better things. [11, p. 12]

A survey of some texts and monographs on special functions exhibits a wide range of responses to the question: What makes a function special? For instance, continuing in the pragmatic vein, Nico Temme says

[u]sually we call a function “special” when the function, just as the logarithm, the exponential and trigonometric functions (the elementary transcendental functions), belongs to the toolbox of the applied mathematician, the physicist, or engineer. Usually there is a particular notation, and a number of properties of the function are known. [26, p. xi]

Somewhat more restrictedly, I. N. Sneddon takes special functions to “arise in the solution of partial differential equations governing the behavior of certain physical quantities.” [21, p. 1] After a very brief introduction to the Laplace equation $\nabla^2\psi = 0$ and a brief discussion of typical solutions, Sneddon remarks that “[i]t is the study of differential equations of this kind which leads to the special functions of mathematical physics. The adjective ‘special’ is used in this connection because we are not, as in analysis, concerned with the general properties of functions, but only with properties of functions which arise in the solution of special problems.” [21, p. 2] This sense of “special” is also somewhat pragmatic—being tied to particular problems of physical interest.

However, other authors are less willing to account for the special nature of special functions *solely or simply* by an appeal to their general usefulness, their usefulness for certain special problems in certain

situations, or because they allow for a shared common currency or culture. For example, James D. Talman, proposes to demonstrate how a wide class of special functions are “matrix elements, or are simply related to matrix elements, of the representations of elementary groups such as rotation groups and Euclidean groups. Many properties of the special functions can then be derived from a unified point of view from the group representation property.”² [25, p. 1] Talman’s conclusion is that

[t]he group theoretic treatment shows that the special functions are special only in that they are related to specific groups. The usefulness of group representation theory for the solution of a variety of physical problems makes it natural that representation matrix elements are important special functions for many problems in mathematical physics. It may further be true that the properties that can be derived group theoretically are their most important ones, since they originate from the “geometric” properties of the functions. [25, p. 2]

Talman’s work was inspired by lectures by Eugene P. Wigner who states in the introduction to Talman’s book that

... the common point of view from which the special functions are here considered, and also the natural classification of their properties, destroys some of the mystique which has surrounded, and still surrounds, these functions. Whether this is a loss or a gain remains for the reader to decide. [25, p. xii]

Clifford Truesdell’s (1948) monograph aims to “provide a general theory which motivates, discovers, and coordinates [certain] seemingly unconnected relations among familiar special functions [here he refers to 35 formulas expressing the relations] ...” [27, p. 7] It predates most such attempts; and focuses, as well, on a particular class of special functions—the “familiar” ones. His work “observes and investigates a hitherto largely neglected property common to the transcendents occurring most frequently in mathematical physics ...” [27, p. 8] This class of *familiar* special functions include Bessel, Legendre, Laguerre, and Hermite functions which have the following “major formal

²Talman does admit that this group theoretic treatment leaves out a wide class of special functions—those for which no group theoretic basis has been found.

properties” in common:

(1) They satisfy ordinary linear differential equations of second order; (2) they satisfy ordinary linear difference equations of second order; (3) with suitable weight functions they form complete sets of orthogonal functions over suitable intervals; (4) they satisfy linear differential-difference [recurrence] relations. The first three of these properties after very long and thorough investigations by numerous excellent mathematicians have yielded but slight clues [sic] to the discovery of such relations as [those 35 mentioned above]. *Our confidence in the essential beauty and perfection of classical analysis would be shaken if in fact these formulas were, so to speak, random effusions of the Divine Mathematician, disjoint and chaotic, so we are driven to conclude that they are consequences of some of the differential recurrence relations* [here follows a list of twelve differential recurrence relations for Bessel, Legendre, Laguerre, and Hermite functions.]. [27, pp. 8–9; My emphasis.]

Truesdell seeks to characterize the special nature of this large class of special functions by demonstrating that they all satisfy a single functional differential recurrence equation that he calls the “F-equation”:

$$\frac{\partial}{\partial z} F(z, \alpha) = F(z, \alpha + 1).$$

Upon examining the group theoretic method for treating special functions initiated by Wigner and carried out by Talman, one sees that the Talman/Wigner method is actually quite similar to Truesdell’s. Talman too employs differential difference equations having the form of the F-equation. The main difference is the group theoretic “spin” (in the political sense) given to the various unifications and classifications.

Truesdell’s insistence that special functions not be chaotic and disjoint effusions of the Divine Mathematician displays somewhat less of a pragmatic answer to the question of the specialness of special functions than do the others. Likewise, the group theoretic approach of Talman and Wigner also seeks a more objective answer to the question. Both of these (related) approaches aim to answer the question by presenting a unifying classification scheme—a scheme that shows what deep *mathematical* properties the different functions share despite obvious “surface” differences.

Surely part of an answer to the question of what makes special functions special will be a well-motivated scheme that groups the different functions together according to shared mathematical features. But does a classificatory schema, by itself, address the truly interesting aspects of special nature of special functions? Don't we want to know why these different functions are so useful? Don't we want to know why the same functions appear over and over again as solutions to disparate problems in the sciences—from fundamental physics to chemistry, biology, economics, etc.?

A natural first response is that what makes them special in this regard is the fact that the differential equations governing various phenomena in these diverse areas are all of the same few general types. Therefore, the solutions will all be quite similar. So it is no wonder that solutions to these few general types of equations crop up over and over again. I think that this response is fine, up to a point. Nevertheless I think that much more can be said about the specialness of such functions.

2 Idealization

In order to begin to motivate this claim, note that almost all applications of mathematics to physical problems involve idealized models of some kind or other. Idealizations, in effect, allow us in many instances to solve the complicated equations governing the phenomenon of interest. In the philosophical literature, there is a traditional approach to modeling that holds that one should try to find the most accurate and detailed mathematical representation of the problem at hand.³ Once this is accomplished (somehow, there is no recipe), one employs the model to make predictions or to try to explain the behavior of interest. If the model fails to capture adequately those features of the phenomenon one is interested in, then there are a couple of things one can do. For instance, one can try to add more detail to the mathematical representation⁴, or one might try to adjust the parameters already appearing in the model so as to better reflect what is going on. Most crucially, on this point of view, the aim is to try to effect a

³I consider the work of Ronald Laymon as representative of this approach to idealization. See for instance, his [17].

⁴By this I mean, one might include mathematical representations of additional factors that may be relevant for the phenomenon under investigation.

kind of convergence between model and reality. Ultimately, the goal is to arrive at a complete (or true) description of the phenomenon of interest. Thus, on this view, a model is better the more details of the real phenomenon it is actually able to represent mathematically. In effect, the idealizations are introduced only to be removed through further work on those details.

This traditional view of modeling is to be contrasted with an approach (or set of approaches) that holds that a good model is one which does not let a lot of these details get in the way. In many cases the full details will not be needed to characterize the phenomenon of interest, and may, in fact, actually detract from an understanding of that phenomenon. This other (nontraditional) approach requires that one find a minimal model—a model “which most economically caricatures the essential physics.” [16, p. 33] The adding of details with the goal of “improving” the minimal model is self-defeating—such improvements are illusory. (I’ve written about this elsewhere and will not go into too much detail here. [3])

Now, often the type of idealization that leads to tractable mathematical models involves, first, an attempt to nondimensionalize the problem so that one can see what terms are relatively “large” and what terms are relatively “small”. Second, one hopes that it is legitimate to ignore the very small terms. Roughly, this can be done by letting the small terms approach zero or by letting the large terms approach infinity; hence, the process here involves the taking of limits—studying various asymptotic behaviors of the functions of interest. This is not by any means *solely* an exercise in pragmatics: It is not simply a means for finding exactly solvable solutions. In today’s world of extraordinary computing capabilities, this analytical practice continues to play a major role in the investigation of physical phenomena. If all we cared about were correct and accurate numerical predictions, then we would not bother with these analytic investigations.

Furthermore, it is often the case that the *dominant* features of certain phenomenon are most effectively represented by taking limits. For instance, consider the case of shocks.

Let’s say we are interested in understanding the behavior of a gas as it moves through a tube. See figure 1. If a collection of the molecules are given a push (say by blowing into the tube at one end), then they will begin to catch up to those in front resulting in a more densely populated region separating two regions of relatively less molecular density. Across this region, molecules will exchange momentum with

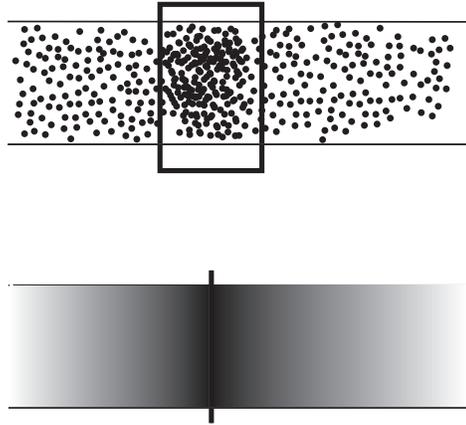


Figure 1: Modelling Shocks

one another as if “some kind of permeable membrane were present.” The region occupied by this “membrane” is a shock. Of course it is very difficult to track the behavior of the individual molecules as they move through the tube and undergo the collisions in the shock region. (This is not to say that computational simulations cannot approximately track such behavior.) But, often the applied mathematician will approach the problem by taking a continuum limit. This is a model in which the collection of molecules in the tube is treated as a continuous fluid. Such a limit will shrink the shock region onto a two dimensional boundary. Upon either side of the boundary, the behavior of the fluid will be governed by the relevant (partial) differential equations of fluid mechanics. However, the behavior across the boundary is not governed by any differential equation at all, but rather by algebraic “jump conditions”—singular behavior across the boundary.

One might think (if you held the more traditional approach to modeling) that the idealization of the collection of molecules to a continuous fluid would make the boundary region unimportant to the physics. After all, the boundary shrinks to two dimensions and is not “law governed.” (All those ignored molecular details ought to be put back in!) Traditional (covering law) accounts of explanation hold that laws do the essential explanatory work, and initial conditions and boundary conditions are given a sort of secondary status. Further, as the boundary is a place where the laws apparently break down, how can the boundary function in a covering *law* explanation?

Mark Wilson has argued that this view—the view that the boundary becomes unimportant to the physics—is mistaken. In fact, the boundary is the most important feature when it come to understanding the behavior of interest. As Wilson notes “the allegedly ‘suppressed details’ have become crushed into a singular (hence not law-governed) factor that still *dominates* the overall behavior through the way in which it constrains the manner in which the ‘law governed regions’ piece together.” [30] The idea is that such boundaries dominate the physics and that often the mathematical modeler’s search focuses on those features to explain what is going on. The limits often yield boundaries that shape or constrain the phenomena. And, it is the elucidation of these shapes that is important for understanding.

In the next section I consider a more detailed example. However, before getting to this, let me try to make a few constructive remarks about how to understand this talk of “dominant features of phenomena.”

It is an incontrovertible fact that nature presents us with patterns and regularities. And, much of scientific theorizing involves trying to understand how these regularities arise. This is not to say that every pattern we observe reflects a genuine lawful feature of the world. Humans are all too ready to see patterns in just about anything.⁵ Neither is it to say that we are interested *only* in investigating “real” regularities and patterns. *Sui generis* phenomena are, of course, also worthy of investigation. As an example of the latter one might think of studying the nature of the transient behavior in a particular electrical circuit before it settles down to a steady state.

Nevertheless, most often it seems that our attention is captured by regularities—repeatable phenomena. It is, in part, the repeatability of phenomena that makes it dominant. That is to say, the repeatability is a *salient* feature that itself leads us to ask about what is responsible for that repeatability. When we couple this feature—the salience of the phenomenon—with the fact (mentioned above) that for all but the simplest empirical generalizations we need to idealize so as to find a adequate mathematical representation, we gain a fuller understanding of the meaning of “dominant feature.” The goal of mathematical modeling is to capture these salient features of the regularity in a mathematical formula. The repeatability of the phenomenon places

⁵Terrence Fine, in his excellent discussion of computational complexity, randomness, and probability, puts the point as follows: “Too keen an eye for pattern will find it anywhere.” [15, p. 120]

a constraint on the nature of the mathematical model. The model must be sufficiently robust or stable under certain kinds of changes to reflect the fact that the phenomenon is repeatable in various situations where many details have changed. The world is constantly changing in myriads of ways, nevertheless, we see the same patterns over and over again in different situations. Idealizing is a means for focusing on exactly those features that are constitutive of the regularity—those features that we see repeated at different times and in different places. Equivalently, idealization, understood in this way, is most broadly seen as a means for removing details that distract from such a focus—those details that can change without affecting the dominant, repeatable behavior of interest.

The mathematical operation that represents the removal of such irrelevant details involves the taking of limits. For instance, in the example just considered, the continuum limit provides a means for ignoring details about molecular interactions in the development of shocks. Most importantly, the taking of limits in this way often imposes mathematical constraints on the equations or formulas that represent the phenomenon of interest. In particular, as just noted, it requires our models to exhibit the appropriate kind of stability under perturbation of various details. Thus our attempt to represent the dominant features of the phenomenon—genuine features of the world—dictates to some extent the nature of the appropriate mathematical representation. That representation, in turn, leads us to investigate in detail, the nature of the imposed constraints. It turns out that in many instances such investigations lead to the discovery of singularities—or places of mathematical “breakdown.” The example of shocks is just one such instance.

The next example illustrates further this interplay between the dominant features of a phenomenon, their mathematical representation, and the constraints imposed upon the representation by our attempt to focus upon those dominant features.

3 Airy’s Equation

In 1838 George Biddell Airy [1] derived a definite integral from the undulatory or wave theory of light from which one can determine the variation in intensity of light near a caustic of geometrical or ray optics. The term “caustic” literally means “burning surface” and in

nature they are extremely bright lines and surfaces caused by the natural focusing of light. They are, in a very obvious sense “dominant” features of light phenomena.

In ray optics, a caustic is an envelope of a family of rays. Unlike the focal point of a lens which is destroyed if the shape of the lens is even slightly perturbed, many caustics of interest are stable against such small perturbations.⁶ The simplest example is called a “fold” caustic, and the ray theory account of the rainbow has it that the main intense bow is located on a fold caustic. See figure 2. One can see that on the “lit” side, there are two rays through every point, while on the dark side of the bow no light rays exist.

The ray theory is capable of locating (only approximately, as it turns out) the primary bow of a rainbow—the caustic surface—but it fails to adequately account for the intensities that one observes. In fact, on the ray theory a caustic is a line on which the intensity of light is strictly speaking infinite. This represents a kind of discontinuity or singularity in the theory. In addition the ray theory is incapable of describing the so-called supernumerary bows—bows resulting from wave interference—that are sometimes seen as faint arcs of alternating pink and green on the lit side of the main rainbow arc.

Airy recognized that the singularity at the caustic was an artifact of the ray theory, and by properly incorporating effects of diffraction using the wave theory, he was able to derive the definite integral of equation (1).

$$Ai(x) \equiv \frac{1}{\pi} \int_0^{\infty} \cos\left(\frac{t^3}{3} + xt\right) dt. \quad (1)$$

The square of this, $|Ai(x)|^2$, gives the intensities of the various supernumerary bows of the rainbow caustic and properly represents the smoothing of the ray theory singularity. [1, p. 1838]

Figure 3 [10, p. 1185] gives an idea of how well the square of $Ai(x)$ can locate and describe the intensities of the wave patterns (the supernumerary bows) decorating the rainbow caustic.

Using a convergent series representation of this function, Airy was able to compute values for $Ai(x)$ for x between -5.6 to +5.6, where x is the distance along a normal to the caustic, negative values on the lit side, positive values in the shadow. Airy’s series converges for all

⁶The mathematical demonstration of this stability involves investigating various structures in the ray theory. Thus, this involves examining the wave theory in the shortwave $\lambda \rightarrow 0$ limit.

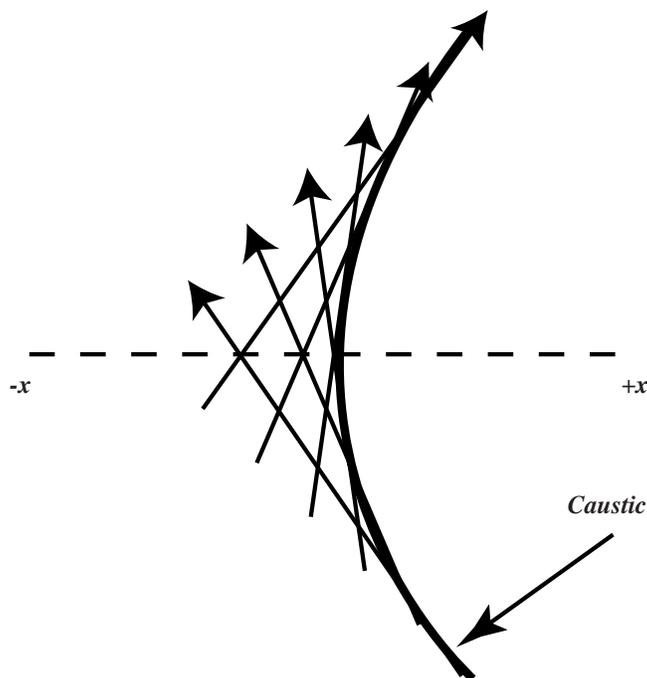


Figure 2: Fold Caustic

values of x and, as a result of this convergence, we may take it to be an exact representation of the Airy function. Unfortunately, this range of values allows one to locate only the first two dark bands on the lit side of the caustic. Because of the slow convergence of the series Airy was unable to extend his calculations beyond this limited range.

George Gabriel Stokes was intrigued by this problem. In 1850, in the first of several discussions having their genesis in Airy’s caustic paper, he noted that despite the fact that Airy’s series converges, “. . . when $[x]$ is at all large the calculation becomes exceeding laborious.” [24, p. 329] Stokes was motivated, therefore, to express the Airy integral as a solution of a differential equation in a form that would exhibit “what terms are of most importance when x is large . . .” [24, p. 331] On the surface it appears that Stokes’ motivation for searching for this form of an asymptotic solution—one which turned out to be what he called “semi-convergent” or divergent—was primarily pragmatic in nature. The convergent expansion was just too unwieldy to use. Nevertheless, I believe that the end results of Stokes’ efforts have

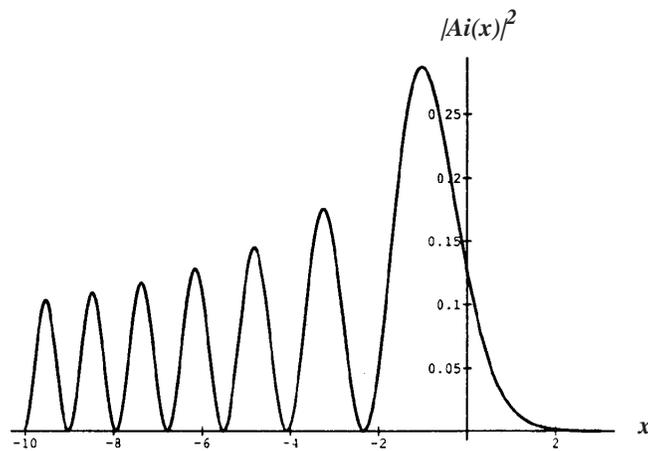
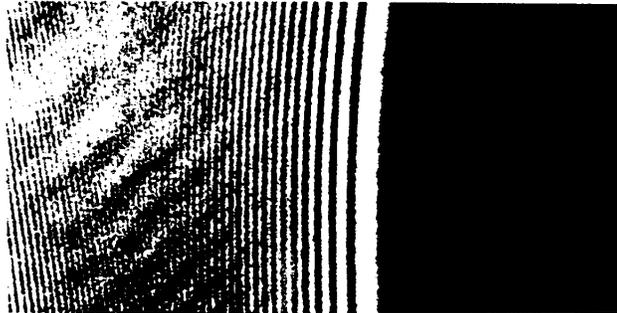


Figure 3: Airy Function and Supernumerary Bows

more import than this simple pragmatic motivation suggests. He was motivated largely by a desire for physical understanding—a desire to display in the most perspicuous fashion *those structures or features that dominate the phenomenon of interest*; namely, the rainbow.

4 Stokes on the Airy Integral

Let me, therefore, discuss Stokes' work and his motivations in a bit more detail. As just noted, his motivation appears at first blush to be largely pragmatic in nature: The convergent series expansion, while exact for all values (real and complex) and, hence, definitive of the function, does not lend itself to easy calculation when the value of its

argument is at all large. As we've seen Airy was able to calculate the locations of only the “first two dark bands in a system of spurious rainbows, whereas Professor Miller was able to observe 30 of these bands.” [24, p. 329] Stokes writes:

After many trials I at last succeeded in putting Mr Airy's integral under a form from which its numerical value can be calculated with extreme facility when $[x]$ [the distance from the geometrical caustic] is large, whether positive or negative, or even moderately large. [24, p. 330]

The next sentence, however, is even more interesting: “Moreover the form of the expression points out, without any numerical calculation the law of the progress of the function when $[x]$ is large.” [24, p. 330] This suggests that there is more to Stokes' solution than simply ease of calculation. Had Stokes had large supercomputers at his disposal, there still would be compelling reasons for engaging in this asymptotic investigation—namely, that such investigations can often highlight important mathematical structures that are hidden in (or obscured by) the exact, convergent expansion.

Stokes derived his asymptotic (divergent) expansion by finding the differential equation for which the Airy integral (1) was a solution and by using what is now known as the WKB method. He examined the general case where the argument can be complex and arrived at the following differential equation:

$$\frac{d^2u}{dz^2} - zu = 0. \quad (2)$$

Stokes reasoned as follows (and, at base, this is the type of asymptotic reasoning behind the WKB method): Let us focus on large values of $|z|$ which reflects our interest in being able to describe the locations and intensities of the bows relatively far from the caustic. If we increase $|z|$ by a small increment δz , the proportionate increase of $|z|$ will be small. That is, for large $|z|$ we can effectively regard z as a constant. If we make this assumption, then equation (2) has the approximate solution:

$$u(z) \equiv Ai(z) \approx Ae^{-\frac{2}{3}z^{\frac{3}{2}}} + Be^{\frac{2}{3}z^{\frac{3}{2}}}, \quad (3)$$

where A and B are arbitrary constants. Stokes noted immediately that “[t]he approximate integral [(3)] points out the existence of circular functions . . . in the true integral.” [24, pp. 334–335] In fact, the

approximate “solution” (3) shows that when z is real and positive, the behavior of the solution to equation (2) can be exponential in character and when z is real and negative, its behavior will be oscillatory or trigonometric in character—whence the expression “circular functions in the true integral.” To get the complete asymptotic expansion, Stokes multiplied the two exponentials (or circular functions) by a series in descending powers of z and solved for various constants. In fact, Stokes discovered that there were two such asymptotic expansions that satisfy the differential equation (2) valid in different sectors of the complex plane. This posed a problem which occupied him at various times throughout the rest of his life. This was to try and understand what is now known as the “Stokes Phenomenon.” It is essentially the question of how the two distinct series are related to one another. I’ve written about this elsewhere [2]. The details are, tangentially relevant to the current discussion of the specialness of special functions.

It is important to understand that Stokes’ approximation argument was guided mainly by physical intuitions about the nature of the phenomenon. In coming up with this alternative series representation of the Airy integral he, in effect, was trying to display in a mathematically perspicuous fashion the important features of the rainbow at distances *relatively* “far” from the ray-theoretic singularity—the caustic.⁷

It is instructive to note the manifest difference between Airy’s convergent expansion and Stokes’ asymptotic, divergent expansion. The leading terms of Stokes’ series describe the salient features of the rainbow—features which Airy’s convergent series does not reveal. The information in the convergent series representation which, recall, is definitive of the function, is decoded only through lengthy numerical calculations and is not manifest in the *form* of the terms in the series. The difference is obvious when one examines the two series “side by side.” Here is Airy’s convergent series:

$$\begin{aligned}
 Ai(z) = & A \left\{ 1 + \frac{9z + 3}{2 \cdot 3} + \frac{9^2 z^6}{2 \cdot 3 \cdot 5 \cdot 6} + \frac{9^3 z^9}{2 \cdot 3 \cdot 5 \cdot 6 \cdot 8 \cdot 9} + \cdots \right\} \\
 & + B \left\{ z + \frac{9z^4}{3 \cdot 4} + \frac{9^2 z^7}{3 \cdot 4 \cdot 6 \cdot 7} + \frac{9^3 z^{10}}{3 \cdot 4 \cdot 6 \cdot 7 \cdot 9 \cdot 10} + \cdots \right\} \quad (4)
 \end{aligned}$$

⁷Recall that Airy was able to locate only the first two bows—those closest to the caustic.

Here is Stokes' series:

$$\begin{aligned}
 Ai(z) \equiv & Cz^{-\frac{1}{4}}e^{-2z^{\frac{3}{2}}} \left\{ 1 - \frac{1 \cdot 5}{1 \cdot 144z^{\frac{3}{2}}} + \frac{1 \cdot 5 \cdot 7 \cdot 11}{1 \cdot 2 \cdot 144^2z^3} - \dots \right\} \\
 & + Dz^{-\frac{1}{4}}e^{2z^{\frac{3}{2}}} \left\{ 1 + \frac{1 \cdot 5}{1 \cdot 144z^{\frac{3}{2}}} + \frac{1 \cdot 5 \cdot 7 \cdot 11}{1 \cdot 2 \cdot 144^2z^3} + \dots \right\} \quad (5)
 \end{aligned}$$

Note how the first terms in Airy's series are the constant A and the expression Bz . These terms tell us virtually nothing about important features of the function. The first terms of Stokes' series, on the other hand, convey much more information. In fact, as we have already seen in equation (3), the first two terms of Stokes' series tell us much about the oscillatory and exponential aspects of the rainbow.

That the "solution," (5), and in particular, its leading terms, explicitly exhibits the oscillatory and exponential character of the "true integral," is a crucial component of our understanding of the physical phenomenon of interest. Our understanding of the patterns present in the rainbow is provided by the relatively *transparent* mathematical representation of these dominant characteristics.

In contrast, while Airy's convergent series provides an "exact" solution to equation (2) for all values of $|z|$, virtually no information about the dominant physical features of the phenomenon is conveyed by the analytical form of the terms of the series. Of course, this is why the convergent series is practically worthless for calculating the locations of the bows in the rainbow for all but those very near to the caustic.

Let me briefly comment upon the similarities between this detailed discussion of the work of Airy and Stokes on trying to understand certain features of the rainbow and the earlier, less detailed, discussion of shocks. The spacings of the light and dark bands in the rainbow and the intensities of these bows are universal phenomena. That is to say they are the same in all rainbows, no matter how the rainbows are formed.⁸ Thus, these are stable dominant features of the rainbow phenomenon and that is one main reason for our studying them in the first place. Our understanding of these universal features requires an investigation into the asymptotics (the limit of short wavelength) of the Airy function as it is only in this asymptotic regime that we can explain why these properties are universal. I've discussed this at length elsewhere [6, 4], but the idea is that the universality in the

⁸For a discussion of this notion of universality, see [6].

behavior of rainbows depends upon the stability of the fold caustic (a mathematical structure) under perturbation. The caustic structure is, in effect, a boundary between two theories—the wave theory and the ray theory of light—and plays a constraining role very much like that played by the two dimensional boundary in the shock problem of the last section. In that example the boundary emerged when we took the continuum limit. In this example, idealized caustic emerges when we take the short wave limit. In both situations, the limits focus our attention on the crucial explanatory features at play, and allow us to ignore countless other physical details that serve only to obscure the most important aspects of the situations.⁹

5 Asymptotic Expansions and Special Functions

Let us now return to the discussion of the nature of special functions.

Note first that the Airy function is a special function and has wide ranging applicability. That is to say it crops up in many diverse problems of applied mathematics: It appears in problems of optics and electromagnetism (as we've just seen), in fluid dynamics, for instance in the description and characterization of solitons, in barrier penetration (tunnelling) problems in quantum mechanics, etc.¹⁰ It is, therefore, an extremely useful function—one whose analytical properties are well-worth knowing.

I think it is fair to say that the standard view of asymptotic representations of various functions is that they are particularly useful. They provide quite accurate numerical values even when one considers very few terms in the series. For example, most applications of the WKB method retain only the first term in the asymptotic expansion. On the other hand, because the late terms of such series typically diverge, they have historically been taken to be inherently

⁹To add just a bit more detail note that Stokes' divergent series representation is accurate for large values of $|z|$; that is, it describes the behavior of light relatively far from the caustic of geometrical optics. Near the caustic (at $z = 0$) the representation breaks down. Nevertheless, the understanding of the universal nature of the patterns of light in a rainbow requires the asymptotic investigation of the light in the neighborhood of the caustic. The singularity—the boundary between wave and ray optics—constrains the solutions away from the singularity on either side.

¹⁰See [28] for an extensive discussion of the Airy function and its applications.

vague and without any coherent meaning. Contemporary understanding of asymptotic expansions rejects this skeptical assessment of the meaningfulness of asymptotic expansions, and, in fact, one can see, already in Stokes' own work, the seeds of this modern point of view.

One particular remark of Stokes' indicates his uneasiness with dismissing asymptotic, divergent series as meaningless despite their extreme pragmatic usefulness. In discussing (what is now called) the Stokes Phenomenon—the discontinuous change in expression (in different phase sectors) that accompanies the asymptotic representation of a continuous function—Stokes notes that “[a] semi-convergent [divergent] series (considered numerically, and apart from its analytical form) defines a function only subject to a certain amount of vagueness . . .” [23, p. 285] Stokes' parenthetical remark here seems to say that while there is numerical vagueness in asymptotic expansions, there may be exact formal information captured in the full asymptotic expansion. This view is, in effect, the starting point for R. B. Dingle's [14] important investigation into the derivation and interpretation of asymptotic expansions.

Dingle observes that

asymptotic expansions are normal, immediately comprehensible, functions of their variables *in so far as functional form is concerned*. The dissimilarity with convergent expansions lies in the impossibility of assigning a precise *numerical* value to the sum (except in the infinite limit) before interpretation of the divergent sequence of late terms. A complete asymptotic expansion of a function $f(x)$ may therefore be defined as an expansion containing an asymptotic series which formally exactly obeys—throughout a certain phase sector—all those relations satisfied by $f(x)$ which do not involve any numerical value of x other than on the infinite circle $|x| \rightarrow \infty$: for instance,

- (i) Functional form as $|x| \rightarrow \infty$, i.e. boundary conditions on f and its derivatives at infinity.
- (ii) Differential, difference and integral equations.
- (iii) Relations involving other parameters incorporated, such as recurrence relations between orders. [14, pp. 19–20]

Dingle calls this observation “the ‘non-numerical compliance’ definition of a complete asymptotic expansion.” [14, p. 19] In maintaining that a complete asymptotic expansion for a given function $f(z)$ is

not inherently vague or meaningless because of the divergence of the asymptotic series in the expansion, Dingle commits himself to providing *an interpretation of the late terms of the series*. In general an asymptotic power series $\sum a_n/z^n$ appearing in a complete expansion typically exhibits the following behavior: For fixed large value of z the series converges rapidly for a finite number of terms N , the error after N terms being less than the $(N + 1)^{\text{st}}$ term. However, despite this initial convergence, the series ultimately diverges. (This is why Stokes calls such series “semi-convergent” and it is also why they are so useful for numerical calculations.) The so-called “late terms” of the series are those that are left out in such an “optimal” truncation. As I said, Dingle is committed to providing an interpretation of these late terms. Such an interpretation removes the aura of vagueness or lack of meaning that traditionally surrounded the use of asymptotic/divergent expansions.

Dingle’s interpretive scheme is based on a theorem of Darboux from 1878. Suppose a function $f(z)$ is singular at some set of points $\{z_i\}$.¹¹ Then the radius of convergence of the power series for $f(z)$ is $\rho = \min|z_i|$. Supposing that $f(z)$ converges within some circle and so can be expanded within that circle as a power series $\sum_0^\infty a_n z^n$. Let z_i be a branch point or pole on or outside this circle. Then near z_i one can write $f(z) = (z_i - z)^{-p_i} f_i(z)$. Here p_i is a positive integer if z_i is a pole or p_i is fractional if z_i is a branch point, and $f_i(z)$ can be written as a Taylor series in $z_i - z$:

$$f_i(z) = f_i(z_i) - (z_i - z)f'_i(z_i) + (z_i - z)^2 f''_i(z_i)/2! - \dots$$

Given this Dingle [14, pp. 140–141] shows that the coefficients a_n of z^n for $n \gg p_i$ will be proportional to

$$(n + p_i - 1)!/n!(p_i - 1)!z_i^{n+p_i}.^{12}$$

Darboux’s theorem shows that the late terms in a convergent Taylor expansion depend only on the behavior of the function in the immediate neighborhood of the singularity closest to the point of expansion. Dingle showed that this fact extends to asymptotic (that is, divergent) series as well. Therefore, almost without exception the late terms of divergent asymptotic series will have the same form. In fact, the late terms ($n \gg 1$)

¹¹This means that at the z_i ’s, f has a pole or branch point.

¹²Note that this has the general form of $(n + \text{constant})! / (\text{variable})^n$.

of *any* asymptotic power series will transpire to be expressible in a standard limiting form $(n + \text{constant})! / (\text{variable})^n$, the accuracy of this limiting representation increasing with n This conclusion . . . is critically important in two ways: first, because it provides a valuable lead on how asymptotic power series and expansions containing them might best be defined; and second, because it shows that substantially a single theory of interpretation will apply equally to late terms of all such asymptotic series. [14, p. 4, with a slight change in notation.]

To illustrate further what John Boyd calls “Darboux’s Principle”¹³, consider the class of functions

$$f(\epsilon) \equiv \int_0^\infty \exp(-t)\Phi(\epsilon t)dt, \quad (6)$$

where $\Phi(z)$ has the power series expansion

$$\Phi(z) = \sum_{n=0}^{\infty} b_n z^n. \quad (7)$$

It follows that

$$f(\epsilon) \sim \sum_{n=0}^{\infty} a_n \epsilon^n; \quad (8)$$

where $a_n = n!b_n$. [13, pp. 29–30]

Boyd points out that since the coefficients, a_n , of the divergent expansion (8) are just those of the power series Φ (equation (7)) times $n!$, it has to be the case that those coefficients are controlled by the singularities of the function $\Phi(z)$

as surely as those $[b_n]$ of the power series [(7)] of Φ itself. In particular, *the singularity of the integrand which is closest to $t = 0$ must determine the leading order of the coefficients of the divergent expansion.* This implies that all $f(\epsilon)$ that have a function $\Phi(z)$ with a convergence-limiting singularity of a given type (pole, square root, etc.) and a given strength (the constant multiplying the singularity) at

¹³This is the principle that the asymptotic expansion in degree n for the coefficients a_n of a series may be derived solely from knowledge of the singularities of the function $f(z)$ that the series represents.

a given point z_c will have coefficients that asymptote to a common form, even if the functions in this class are wildly different otherwise. [13, p. 30]

Dingle's results show that it really almost does not matter what the nature of the singularity of the function is—whether it is a pole, a branch point, a logarithm, etc. Nor does it matter whether the asymptotic expansion derives from an integral representation, the WKB method, from a differential equation, etc. In virtually all cases, the late terms in the asymptotic expansions will have one of four basic forms that he calls “terminants.” Essentially, Dingle's program (further developed by others including Berry [9, 8], Olver [19], Olde Daalhuis [18], et al.) explicitly shows how a vast range of functions—including a majority of the so-called special functions—are such that they can be transformed through asymptotic analysis so as to exhibit a common or *universal* pattern; namely,

- Function = first n terms of asymptotic series + n^{th} × terminant.[14, p. 411]

As Dingle says, “it is the Darboux theorem which lies at the heart of the reasons *why* this common pattern exists.” [14, p. 411]

Dingle's book provides evidence of just how general this method (and its further elaboration by others) really is. He shows how to derive asymptotic expansions for many different functions, including those definable by convergent expansions such as the incomplete factorial function, the exponential integral, the confluent hypergeometric function, the Fermi-Dirac integral, spherical Bessel functions, Hermite polynomials, Legendre polynomials, etc. He shows how to derive asymptotic expansions for other (special) functions that arise from various integral representations related to the Laplace integral. He shows how to derive asymptotic expansions for solutions to various types of homogeneous differential equations, and for solutions to various types of inhomogeneous differential equations, etc. The point is that the range of problems and functions that can be given asymptotic representations is extremely broad; yet, all of these representations require to the same method of asymptotic interpretation that we have discussed in this section.

John Boyd, in the paper cited above, provides a 2 1/2 page table exhibiting the range of problems and functions to which the interpretive scheme initiated by Dingle applies. [13, pp. 42–44] It includes, many problems from physics including (among others) quantum tun-

nelling, problems in nonlinear optics, and critical phenomena. Boyd states that “Dingle’s ideas of generic forms for the late terms in asymptotic series and universal terminants now seem as important to the rise of exponential asymptotics as the comet-crash(?) that put an end to the dinosaurs was in biology.” [13, p. 79] But he recognizes that this insight was indeed slow in coming.

The point I would like to emphasize here is that if one examines special functions from an asymptotic point of view—looking for representations of the functions in terms of *formally exact* asymptotic expansions—one encounters a deep universal classificatory scheme. In effect, it is the universality of the terminants—the fact that the late terms in virtually all such expansions behave the same way—that can be taken as indicative of the specialness of the functions. Local facts about singularities of the different functions have global, that is, universal consequences.

6 Special Functions and the World

The discussion so far has led us to a different, and perhaps, better mathematical classification of a very broad class of special functions. It is a classification that takes *the formally exact asymptotic representations of the functions as basic* and unifies the class through the common interpretation of the late terms in the expansions of the special functions. But how does this help us answer the questions posed above in section 1. Recall that these were the following: Does a classificatory schema, by itself, address the truly interesting aspects of special nature of special functions? Don’t we want to know why these different functions are so useful? Don’t we want to know why the same functions appear over and over again as solutions to disparate problems in the sciences—from fundamental physics to chemistry, biology, economics, etc.?

Now Berry provided a somewhat rhetorical answer to such questions about the specialness of special functions. Recall that he says: “What are they [special functions], other than just names for mathematical objects that are useful only in situations of contrived simplicity?” [11, p.12] In asking this question, he appears to be asserting that what makes them special (in addition to any classificatory scheme based upon common mathematical properties) is their usefulness for understanding simplified situations in the world. To a certain extent I

think this is correct. However, the *simplified situations* in which special functions are useful are *far from contrived*. I have tried to argue here (and, in more detail, elsewhere) that most applications of mathematics to the description and explanation of physical phenomena do involve simplifications of a sort. They involve, almost always, idealizations in which certain parameters take on a limiting value of zero or infinity. They involve, that is to say, asymptotics. Furthermore, I believe, in many instances such limiting idealizations are *essential* for obtaining genuine scientific explanation and understanding. [5, 7]

Given this it seems reasonable to maintain that many special functions are special for more than simple pragmatic reasons. They are not special simply because they appear in the physicist’s, applied mathematician’s, and engineer’s toolboxes. Furthermore, special functions are not special simply because they share some deep mathematical properties. Recall this is the point of view of Truesdell and of Talman/Wigner. On their proposals, what makes some functions special is that despite “surface” differences, they are each solutions to the “F-equation” (for Truesdell), or they possess similar group representations (for Talman and Wigner). While these classificatory schemes suffice to bring some order to the effusions of the Divine Mathematician, they do not fully capture the special nature of the special functions.¹⁴

From the point of view presented here, the shared mathematical features that serve to unify the special functions—the universal form of their asymptotic expansions—depend upon *certain features of the world*. What Truesdell, Talman and many others miss is how the world informs and determines the relevant mathematical properties that unify the diverse special functions.

As I noted, in many investigations of physical phenomena we find *dominant* physical features—those features that constrain or shape the phenomena. These are things like shocks and the highly intense light appearing in the neighborhood of ray theoretic caustics. They are features that are most effectively modeled by taking limits.¹⁵

¹⁴In fact, the schemes of Truesdell (and of Talman and Wigner) will not be as broadly unifying as the asymptotic unification provided by Dingle, Berry, and others. For instance, the Truesdell scheme does not include so-called higher order diffraction catastrophe integrals—a sequence of special functions of which the Airy integral is just the first order. (Michael Berry, private communication.) For a discussion of this hierarchy see [12].

¹⁵Note that such dominant features need not possess robust observability. Tunnelling phenomena in quantum mechanics, for example, also require asymptotics for their de-

Limiting idealizations are most effective for examining what goes on at places where the “laws” break down—that is, at places of singularities in the governing equations of the phenomena. These “physical” singularities and their “effects”—how they dominate the observed phenomena—are themselves best investigated through asymptotic representations of the solutions to the relevant governing equations. The example of the Airy integral is a case in point. By using Stokes’ asymptotic representation we get superb representations of the nature of the diffraction relatively far (large $|z|$) from the dominating “physical” singularity—the caustic.

Here one can easily see the interplay between the “physical” singularities and the mathematical singularities. The ray-theoretic singularities need to be dealt with by the wave theory—this is just what Airy did in deriving his integral equation. However, as I noted, this ray-theoretic singularity dominates the physical phenomenon. As a result, the *mathematical* singularities in the representative asymptotic equations constrain the structure of the solutions to the governing equations and their asymptotics provides the unification necessary to give an interesting answer to the question of why those solutions—the special functions—are special.

A clear statement of this point of view can be found in the Foreword to a recent discussion of special functions and singularities. Alfred Seeger describes the connections as follows:

In mapping a complex physical situation onto manageable mathematics, location and character of the singularities reflect the essentials of the situation, whereas the parameters not directly associated with the singularities usually carry incidental information, e.g. on the physical properties of the specific material under consideration. Recognizing this led to a new appreciation of the importance of asymptotic expansions and of the Stokes’ phenomenon. On the mathematical side, it is the singularities of the differential equations resulting from the mapping that determine the character of the solutions. [20, p. vii]

The arguments above are meant to support this point of view. I have argued that the combination of the complexity of the physical situation with the fact that, in spite of its complexity, the world presents

scription and explanation. “Dominant” does not necessarily mean “powerful.” As the discussion above tries to show, it means (at least) repeatable and, hence, stable.

us with regular, stable features to be accounted for, requires that we idealize in a certain way. This allows us to focus on what Seeger here calls the “essentials of the situation.” In turn, our attention is focused on certain singularities in the relevant mathematical equations which require asymptotic investigations. This interplay between physics and mathematics—mediated by singularities—has not, in my opinion, been sufficiently appreciated in the philosophical literature.

Further, as I begin to discuss in the next section, this interplay is important for understanding at least one fundamental problem in the philosophy of mathematics.

7 Conclusion: Special Functions and the Philosophy of Mathematics

Therefore, let me make a few comments about how I think this discussion of special functions may shed some light on one of the most difficult problems in the philosophy of mathematics.

The problem is what Wigner famously calls “the unreasonable effectiveness of mathematics in the natural sciences.” He claims that the “appropriateness of the language of mathematics for the formulation of the laws of physics” is a “miracle”—“a wonderful gift which we neither understand nor deserve.” [29, p. 237] The study of the asymptotics of special functions and the reasons for their usefulness in mathematical physics may very well help to dispel the appearance of the miraculous (and for some, the divine).

Mark Steiner in his book, *The Applicability of Mathematics as a Philosophical Problem* [22], argues that contemporary mathematical physicists employ various strategies—particularly, analogical strategies—in forming and discovering new theories about unobservable aspects of the world. He is concerned to show that these analogical strategies are fully anthropocentric—that is, that they depend for their success, upon humans having a special place in the world. Steiner’s arguments involve extended examples in which he tries to show just how disconnected the world really is from the mathematical descriptions we employ to represent it. In large part he aims to maintain the miraculous nature of the applicability of mathematics to the world by demonstrating repeatedly that physicists employ analogies in discovery that are tied primarily to the formalism of existing theories and that cannot in any way be taken to be physically motivated.

If the above account of what makes special functions special provides an answer to the why-questions about the usefulness and ubiquity of such functions, then I think that we can, to some extent, begin to understand the wonderful gift of the applicability of mathematics. We have seen that the very nature of many phenomena that are investigated places constraints on the nature of those investigations. The dominant and real features of phenomena require that we employ limiting idealizations in forming the mathematical equations with which we may represent the phenomena.

Truesdell was concerned to show that the Divine Mathematician does not present us with a set of special functions that fail to be unified in some appropriate way. We need, on his view, to maintain our faith in the “essential beauty and perfection of classical analysis” and this can only happen if there is some nonrandom unification of the special functions. I hope that the discussion here leads us to question the anthropocentric role of the mathematician’s appreciation for beauty (or formal analogy) as an important criterion for what arguably should be paradigm examples of mathematics’ applicability to the world; namely, the special functions.

References

- [1] G. B. Airy. On the intensity of light in the neighborhood of a caustic. *Transactions of the Cambridge Philosophical Society*, 6:379–403, 1838.
- [2] R. W. Batterman. “Into a mist”: Asymptotic theories on a caustic. *Studies in the History and Philosophy of Modern Physics*, 28(3):395–413, 1997.
- [3] R. W. Batterman. Asymptotics and the role of minimal models. *The British Journal for the Philosophy of Science*, 53:21–38, 2002.
- [4] R. W. Batterman. “Response to Belot’s ‘Whose devil? Which details?’ ”. *Philosophy of Science*, 72(1):154–163, 2005.
- [5] R. W. Batterman. ‘Fundamental’ theory: Molecular dynamics vs. hydrodynamics. <http://philsci-archive.pitt.edu/archive/00000582/>, Preprint.

- [6] Robert W. Batterman. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford Studies in Philosophy of Science. Oxford University Press, 2002.
- [7] Robert W. Batterman. Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics*, 36:225–244, 2005.
- [8] M. V. Berry. Asymptotics, superasymptotics, hyperasymptotics. . . . In Harvey Segur, Saleh Tanveer, and Herbert Levine, editors, *Asymptotics Beyond All Orders*, Nato ASI Series, New York, 1991. Plenum Press.
- [9] Michael V. Berry. Stokes’ phenomenon: Smoothing a Victorian discontinuity. *Publ. Math. of the Institut des Hautes Études Scientifique*, 68:211–221, 1989.
- [10] Michael V. Berry. Beyond rainbows. *Current Science*, 59(21 & 22):1175–1191, 1990.
- [11] Michael V. Berry. Why are special functions special? *Physics Today*, pages 11–12, April 2001.
- [12] Michael V. Berry and C. Upstill. Catastrophe optics: Morphologies of caustics and their diffraction patterns. *Progress in Optics*, 18:257–346, 1980.
- [13] John P. Boyd. The devil’s invention: Asymptotic, superasymptotic and hyperasymptotic series. *Acta Applicandae mathematicae*, 56:1–98, 1999.
- [14] R. B. Dingle. *Asymptotic Expansions: Their Derivation and Interpretation*. Academic Press, London, 1973.
- [15] Terrence L. Fine. *Theories of Probability: An Examination of Foundations*. Academic Press, New York, 1973.
- [16] Nigel Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group*. Number 85 in Frontiers in Physics. Addison-Wesley, Reading, Massachusetts, 1992.
- [17] Ronald Laymon. Idealization, explanation, and confirmation. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1980, Volume 1: Contributed Papers, pages 336–350, 1980.
- [18] Adri B. Olde Daalhuis. Exponential asymptotics. In E. Koelink and W. Van Assche, editors, *Orthogonal Polynomials and Special*

- Functions*, volume 1817 of *Lecture Notes in Mathematics*, pages 211–244. Springer-Verlag, Berlin, 2004.
- [19] Frank W. J. Olver. *Asymptotics and Special Functions*. A K P, Wellesley, Massachusetts, 1997.
 - [20] Sergei Yu. Slavyanov and Wolfgang Lay. *Special Functions: A Unified Theory Based on Singularities*. Oxford University Press, Oxford, 2000.
 - [21] Ian N. Sneddon. *Special Functions of Mathematical Physics and Chemistry, Third edition*. Longman, London, 1980.
 - [22] Mark Steiner. *The Applicability of Mathematics as a Philosophical Problem*. Harvard University Press, Cambridge, Massachusetts, 1998.
 - [23] George G. Stokes. On the discontinuity of arbitrary constants that appear as multipliers of semi-convergent series. In *Mathematical and Physical Papers by the Late Sir George Gabriel Stokes*, volume 5, pages 283–287. Johnson Reprint Corporation, 1966.
 - [24] George G. Stokes. On the numerical calculation of a class of definite integrals and infinite series. In *Mathematical and Physical Papers by the Late Sir George Gabriel Stokes*, volume 2, pages 329–357. Johnson Reprint Corporation, 1966.
 - [25] James D. Talman. *Special Functions: A Group Theoretic Approach*. W. A. Benjamin, 1968.
 - [26] Nico M. Temme. *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. John Wiley & Sons, New York, 1996.
 - [27] Clifford Truesdell. *An Essay Toward a Unified Theory of Special Functions based upon the functional equation $\frac{\partial}{\partial z}F(z, \alpha) = F(z, \alpha+1)$* . Number 18 in *Annals of Mathematics Studies*. Princeton University Press, 1948.
 - [28] Olivier Vallée and Manuel Soares. *Airy Functions and Applications to Physics*. Imperial College Press, London, 2004.
 - [29] Eugene P. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. In *Symmetries and Reflections: Scientific Essays of Eugene P. Wigner*. Indiana University Press, 1967.
 - [30] Mark Wilson. Comments on “the devil in the details”. 2003.