

The concept of "character" in Dirichlet's theorem on primes in an arithmetic progression

Author(s): Jeremy Avigad and Rebecca Morris

Source: *Archive for History of Exact Sciences*, Vol. 68, No. 3 (May 2014), pp. 265-326

Published by: Springer

Stable URL: <https://www.jstor.org/stable/24569604>

Accessed: 19-05-2020 09:59 UTC

REFERENCES

Linked references are available on JSTOR for this article:

https://www.jstor.org/stable/24569604?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Archive for History of Exact Sciences*

The concept of “character” in Dirichlet’s theorem on primes in an arithmetic progression

Jeremy Avigad · Rebecca Morris

Received: 22 May 2013 / Published online: 23 July 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract In 1837, Dirichlet proved that there are infinitely many primes in any arithmetic progression in which the terms do not all share a common factor. We survey implicit and explicit uses of *Dirichlet characters* in presentations of Dirichlet’s proof in the nineteenth and early twentieth centuries, with an eye toward understanding some of the pragmatic pressures that shaped the evolution of modern mathematical method.

Contents

1	Introduction	266
2	Functions in the nineteenth century	267
2.1	The generalization of the function concept	267
2.2	The evolution of the function concept	274
3	Dirichlet’s theorem	275

Communicated by: Jeremy Gray.

This essay draws extensively on the second author’s Carnegie Mellon MS thesis (Morris 2011). We are grateful to Michael Detlefsen and the participants in his *Ideals of Proof* workshop, which provided feedback on portions of this material in July 2011, and to Jeremy Gray for helpful comments. Avigad’s work has been partially supported by National Science Foundation grant DMS-1068829 and Air Force Office of Scientific Research grant FA9550-12-1-0370.

J. Avigad (✉)

Department of Philosophy and Department of Mathematical Sciences, Carnegie Mellon University,
Pittsburgh, PA, USA
e-mail: avigad@cmu.edu

R. Morris

Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: rebecca.lea.morris@gmail.com

3.1	Euler's proof that there are infinitely many primes	276
3.2	Group characters	278
3.3	Dirichlet characters and L -series	280
3.4	Contemporary proofs of Dirichlet's theorem	281
4	Modern aspects of contemporary proofs	283
5	Dirichlet's original proof	285
6	Later presentations	291
6.1	Dirichlet	291
6.2	Dedekind	293
6.3	Weber	297
6.4	De la Vallée-Poussin	299
6.5	Hadamard	302
6.6	Kronecker	304
6.7	Landau	309
7	Changes in mathematical method	313
8	Conclusions	321

1 Introduction

Historians commonly take the “modern” age of mathematics to have begun in the nineteenth century. But although there is consensus that the events of that century had a transformational effect on mathematical thought, it is not easy to sum up exactly what changed and why. Aspects of the transformation include an increasingly abstract view of mathematical objects; the rise of algebraic methods; the unification of disparate branches of the subject; evolving standards of rigor in argumentation; a newfound boldness in dealing with the infinite; emphasis on “conceptual” understanding and a concomitant de-emphasis of calculation; the use of (informal) set-theoretic language and methods; and concerns to identify a foundational basis to support the new developments.¹ It is still an important historical and philosophical task to better understand these components, and the complex interactions between them.

A good deal of attention has been given to early appearances of set-theoretic and structural language, including the use of equivalence relations and ideals in algebra; the expansion of the function concept in analysis and its generalization to other mathematical domains; foundational constructions of number systems from the natural numbers to the reals and beyond; and the overall conception of mathematics as the study of structures and spaces, often characterized in set-theoretic terms (Ferreirós 1999). These provide clear and focused manifestations of the changes that took place.

In this essay, we will consider certain functions, known as “Dirichlet characters,” and their role in proving a seminal 1837 theorem due to Dirichlet, which states that there are infinitely many prime numbers in any arithmetic progression in which the terms do not all share a common factor. As far as abstract objects go, characters are fairly benign: For a given positive integer m , there are only finitely many characters modulo m , and each one can be described exactly by giving its value on the finitely

¹ The essays in Ferreirós and Gray (2006) provide an overview.

many residue classes modulo m . Moreover, it is not terribly hard to provide a concrete and exhaustive description of the set of all such characters.

Nonetheless, we will argue that the evolution of the treatment of characters over the course of the nineteenth century illustrates important themes in the overall transformation of mathematical thought. Indeed, we will try to call attention to changes in mathematical method that have by now become so ingrained that it is hard for us today to appreciate their significance. An important feature of contemporary proofs of Dirichlet’s theorem is that they are higher order, which is to say, one treats characters on a par with mathematical objects like the natural numbers: One sums over characters, quantifies over characters, applies functions to characters, forms sets of characters, considers groups whose elements are characters, and so on. This runs counter to an “intensional” view of characters as *expressions* that play a role in ordinary mathematical language that is fundamentally different from that of the natural numbers or the real numbers. Our narrative traces a gradual transition from this latter view to the contemporary one.

In tracing this evolution, we will set aside other important aspects of the history of Dirichlet’s theorem, such as the evolution of number-theoretic and analytic ideas, and the use of analysis in number theory. We will also for the most part set aside other nineteenth-century developments that invoked the use of set-theoretic language. Our hope is that focusing on the notion of character, and its evolving role in a single mathematical proof, will help clarify some of the mathematical pressures that contributed to the adoption of a modern standpoint.

The outline of this paper is as follows. In Sect. 2, we situate our study amid a host of topics related to the development of the function concept, noting that for most of the nineteenth century the word “function” was used exclusively in connection with functions on the real or complex numbers. In Sect. 3, we discuss contemporary presentations of Dirichlet’s proof, and in Sect. 4, we highlight various senses in which these presentations treat characters as objects. In contrast, Sect. 5 describes Dirichlet’s own presentation of his proof, in which the notion of a character does not figure at all. Section 6 then traces a gradual transition, as characters are transformed from shade-like entities in the original proof to the fully embodied objects we take them to be today. Section 7 analyzes the forces that shaped the transition, and Sect. 8 draws some conclusions. In a companion paper to this one Avigad and Morris (n.d.), we explore further philosophical aspects of the treatment of characters as objects in the history of Dirichlet’s theorem.

2 Functions in the nineteenth century

2.1 The generalization of the function concept

Nineteenth-century mathematicians dealt with a number of objects that we, today, view as instances of the function concept. Analysis dealt with functions defined on the continuum, that is, functions from the real numbers, \mathbb{R} , to \mathbb{R} . Cauchy and others extended the subject to include functions from the complex numbers, \mathbb{C} , to \mathbb{C} . Analysts also commonly dealt with sequences and series, which can be viewed as functions from

the natural numbers, \mathbb{N} , to a target domain, such as the integers or the reals. Number theory dealt with functions like the Euler totient function, φ , which maps \mathbb{N} to \mathbb{N} . Geometry dealt with transformations of the Euclidean or projective plane, which can be viewed as invertible maps from the space to itself. Galois' theory of equations focused on substitutions, or permutations, of a finite set A , which is to say bijective functions from A to A . Below we will see how developments in number theory led to the study of characters, that is, functions from the integers, \mathbb{Z} , to \mathbb{C} , or from the group $(\mathbb{Z}/m\mathbb{Z})^*$ of units modulo m to \mathbb{C} . By the end of the century, Dedekind and Cantor considered arbitrary mappings, or correspondences, between domains.

Despite this diversity, most of the literature on the evolution of the function concept (including Chorlay n.d.; Kleiner 1989; Luzin 1998; Monna 1972; Youschkevitch 1976) has focused on functions on the real or complex numbers. Surveys typically trace the evolution of the concept from the introduction of the term "function" ("functio") by Euler (1748) in 1748, to the mention of "arbitrary functions" ("fonction arbitraire") in the title of Dirichlet's seminal paper of 1829 (Dirichlet 1829), to the dramatically opposed treatment of the function concept by Riemann and Weierstrass.² There is a very good reason for this: In the early nineteenth century, the word "function" was used almost exclusively in this narrow sense. Indeed, there is little in the early nineteenth-century literature that suggests any family resemblances between the kinds of mathematical objects enumerated in the last paragraph, let alone subsumption under a single overarching concept.

For example, sequences of real numbers were referred to as sequences or series or often just introduced by displaying the initial elements a_0, a_1, a_2, \dots with an ellipsis at the end. The word "series" ("série" in French, "Reihe" in German) was also used to describe a finite or infinite sum. For example, Dirichlet wrote the following in 1835:

Now let

$$F(\alpha) = b_0 + b_1 \cos \alpha + b_2 \cos 2\alpha + \dots = \sum b_i \cos i\alpha$$

be an arbitrary finite or infinite series, whose coefficients are independent of α .³ (Dirichlet 1835, p. 249)

The dependence of the coefficient b_i on i was indicated by the subscript, just as we do today.

Terminology governing number-theoretic functions was more varied. In the eighteenth century, Euler tended to refer to number-theoretic functions as "symbols" or "characters." When introducing what we now call the totient function in 1781, he wrote:

² See Bottazzini and Gray (2013) for a history of complex function theory, which includes a detailed treatment of the work of Riemann and Weierstrass, in particular.

³ "Es sei jetzt:

$$F(\alpha) = b_0 + b_1 \cos \alpha + b_2 \cos 2\alpha + \dots = \sum b_i \cos i\alpha$$

eine beliebige endliche oder unendliche Reihe, deren Coefficienten von α unabhängig sind."

...let the character πD denote that multitude of numbers which are less than D and which have no common divisor with it.⁴ (Euler 1784, p. 2)

In 1801, in § 38 of the *Disquisitiones Arithmeticae* (Gauss 1801), Gauss introduced the totient function as a notation:

For brevity we will designate the number of positive numbers which are relatively prime to the given number and smaller than it by the prefix φ . We seek therefore φA .⁵

In the following section, he referred to the “character” φ . Later, in § 52, he defined another number-theoretic function, ψ , and in § 53 compared it to the φ “symbol” (“signo”). In the *Disquisitiones*, the word “function” is never used to describe such entities.

Early in the nineteenth century, in the field of projective geometry, Möbius studied general “relations” between figures in the plane, such as “affinities” and “collineations.”⁶ In his *Erlangen Program* of 1872, Klein considered invertible “transformations” (“Transformationen”) of a space, and the group they form under the operation of “composition” (“Zusammensetzung”) (Klein 1893). Once again, in neither case is there any hint that such transformations bear a connection to a more general function concept.

In algebra, Galois’ groups consisted of “substitutions” of the roots. Cayley’s famous 1854 paper, which provided the first axiomatization of the group concept, begins as follows:

Let θ be a symbol or operation, which may, if we please, have for its operand, not a single quantity x , but a system (x, y, \dots) , so that

$$\theta(x, y, \dots) = (x', y', \dots),$$

where x', y', \dots are any functions whatever of x, y, \dots , it is not even necessary that x', y', \dots should be the same in number with x, y, \dots . In particular, x', y', \dots may represent a permutation of x, y, \dots , θ is in this case what is termed a substitution; and if, instead of a set x, y, \dots , the operand is a single quantity x , so that $\theta x = x' = f x$, θ is an ordinary function symbol.⁷ (Cayley 1854, p. 123)

Notice that although we can assume that each quantity x', y', \dots is a function of x, y, \dots , for Cayley, θ is, in general, an operation and not a function. As examples of cases where θ is a function, he singled out multiplication by quaternions and examples arising from the study of elliptic functions. But he makes it clear that it is the operations, more generally, that are assumed to satisfy the associative law, under composition:

⁴ “Quod quo facilius praestari possit, denotet character φD multitudinem istam numerorum ipso D minorum, et qui cum eo nullum habeant divisorem communem.” (Euler 1784, p. 19)

⁵ “Designemus brevitate gratia multitudinem numerorum positivorum ad numerum datum primorum ipsoque minorum per praeifixum characterem φ . Quaeritur itaque φA .”

⁶ See the discussion in Wussing (1984, pp. 35–40).

⁷ This excerpt is quoted and discussed in Pengelley (2005).

A symbol $\theta\varphi$ denotes the compound operation, the performance of which is equivalent to the performance, first of the operation φ , and then of the operation θ ; $\theta\varphi$ is of course in general different from $\varphi\theta$. But the symbols θ, φ, \dots are in general such that $\theta.\varphi\chi = \theta\varphi.\chi$, &c., so that $\theta\varphi\chi, \theta\varphi\chi\omega$, &c. have a definite signification independent of the particular mode of compounding the symbols...

Dedekind's later treatment of Galois theory (Dedekind 1894) focused on the group of automorphisms of a field rather than permutations of roots, but even in this context, he used the term "substitution" for an invertible map from a field to itself, and "permutation" for a substitution that is moreover a homomorphism with respect to the field structure. Once again, there seems to be nothing to link these substitutions and permutations with the function concept from analysis. Even as late as 1895, when Cantor presented his mature theory of the infinite (Cantor 1895), he described a one-to-one correspondence as a "law of association" ("Zuordnungsgesetz") between sets, with nothing to suggest that these correspondences had anything to do with the objects of function theory with which he had begun his mathematical career.

The shift from using the term "function" exclusively for functions defined on the continuum to those defined on more general domains was gradual. We have found a very early instance of the phrase "number-theoretic function" ("zahlentheoretische Funktion") in an 1850 paper by Eisenstein, which begins with a notably self-conscious justification for the use of the term:

Since, with the concept of a function, one moved away from the necessity of having an analytic construction, and began to take its essence to be a tabular collection of values associated to the values of one or several variables, it became possible to take the concept to include functions which, due to conditions of an arithmetic nature, have a determinate sense only when the variables occurring in them have integral values, or only for certain value-combinations arising from the natural number series. For intermediate values, such functions remain indeterminate and arbitrary, or without any meaning.⁸ (Eisenstein 1850, p. 706)

We might have expected Eisenstein to observe, more simply, that once one starts to think of a function as a correspondence between input and output values, it is reasonable to transfer the notion to correspondences between domains other than the real or complex numbers, described by means other than an analytic expression. Instead, he adopted the surprising strategy of viewing number-theoretic functions as *partial functions on the real numbers*. This is a nice illustration of the fact that concepts are often stretched and transformed in surprising ways, a theme that is the central study of Mark Wilson's book, *Wandering Significance* (Wilson 2006). In any case, the passage

⁸ "Seit man bei dem Begriffe der Funktion von der Nothwendigkeit der analytischen Zusammensetzung abgehend, das Wesen derselben in die tabellarische Zusammenstellung einer Reihe von zugehörigen Werthen mit den Werthen des order der (mehrerer) Variabeln zu setzen anfang, war es möglich, auch solche Funktionen unter diesen Begriff mit aufzunehmen, welche aus Bedingungen arithmetischer Natur entspringend nur für ganze Werthe oder nur für gewisse aus der natürlichen Zahlenreihe hervorgehende Werthe und Werth-Combinationen der in ihnen vorkommenden Variabeln einen bestimmten Sinn erhalten, während sie für die Zwischenwerthe entweder unbestimmt und willkürlich oder ohne alle Bedeutung bleiben." We are grateful to Wilfried Sieg for help with the translation.

makes it clear that the modern notion of a function defined on arbitrary domains was far from the mid-nineteenth-century mindset.

We have already noted that in 1854, Cayley considered multiplication by a quaternion as a function. This seems to fit with the view of the natural progression from real numbers to complex numbers to quaternions, as generalizations of the concept of magnitude. One finds another hint of expansion in the title of Dedekind’s 1854 *Habilitation* lecture, “On the introduction of new functions in mathematics” (“Über die Einführung neuer Funktionen in der Mathematik”) (Dedekind 1854). In this lecture, Dedekind discussed the way that the domain of natural numbers was gradually expanded to include the integers, real numbers, and complex numbers, while extending and preserving the properties of familiar operations like addition and division. But the evocative use of the word “function” in the title is tempered by the contents of the lecture itself, where the word “operation” is used exclusively when it comes for functions defined on the natural numbers and integers.

To our knowledge, it is not until 1879 that one finds the word “function” used to describe correspondences between arbitrary domains, and then, in that year, this occurred in two remarkable sources. The first is due to Dedekind. In 1879, in a supplement dealing with quadratic forms in the third edition of his presentation of Dirichlet’s lectures on number theory (Dirichlet 1863), he defined the general notion of a character on a class group (i.e., a function that maps equivalence classes of ideals in an algebraic number field to the complex numbers):

...the function $\chi(a)$ also possesses the property that it takes the same value on all ideals a belonging to the same class A ; this value is therefore appropriately denoted by $\chi(A)$ and is clearly always an h th root of unity. Such functions χ , which in an extended sense can be termed *characters*, always exist; and indeed it follows easily from the theorems mentioned at the conclusion of §149 that the class number h is also the number of all distinct characters $\chi_1, \chi_2, \dots, \chi_h$ and that every class A is completely characterized, i.e. is distinguished from all other classes, by the h values $\chi_1(A), \chi_2(A), \dots, \chi_h(A)$.⁹

In other words, Dedekind referred to a mapping from a certain algebraic structure to the complex numbers as a “function.” There is no reason to think that he meant to limit the terminology to this particular structure, and, indeed, by 1882, Weber (1882) described characters on arbitrary groups as “functions.”¹⁰ In his foundational essay *Was sind und was sollen die Zahlen* (Dedekind 1888) of 1888, Dedekind considered

⁹ “...die Function $\chi(a)$ ausser der Eigenschaft ...noch die andere besitzt, für alle derselben Classe A angehörnden Ideale a denselben Werth anzunehmen, welcher mithin zweckmässig durch $\chi(A)$ bezeichnet wird und offenbar immer eine h^{te} Wurzel der Einheit ist. Solche Functionen χ , die man im erweiterten Sinn *Charaktere* nennen kann, existiren immer, und zwar geht aus den am Schlusse des §149 erwähnten Sätzen leicht hervor, dass die Classenzahl h zugleich die Anzahl aller verschiedenen Charaktere $\chi_1, \chi_2, \dots, \chi_h$ ist, und dass jede Classe A durch die ihr entsprechenden h Werthe $\chi_1(A), \chi_2(A), \dots, \chi_h(A)$ vollständig charakterisirt, d.h. von allen anderen Classen unterschieden wird.” The quotation appears in § 178 in the 1879 edition of the *Vorlesungen* (Dirichlet 1863), and in § 184 of the 1894 edition, which is reproduced in Dedekind’s *Werke* (Dedekind 1932). The translation above is by Hawkins (1971, p. 149).

¹⁰ We discuss Weber’s 1882 paper and provide more background on the history of characters, in Sects. 6.2 and 6.3 below.

arbitrary mappings (“Abbildung”) between domains, but, curiously, seems to distinguish mappings from functions.¹¹

The other landmark source of 1879 is the philosopher Gottlob Frege’s *Begriffsschrift* (Frege 1879). Here Frege defined the notion of a function as follows:

If, in an expression (whose content need not be a judgeable content), a simple or complex symbol occurs in one or more places, and we think of it as replaceable at all or some of its occurrences by another symbol (but everywhere by the same symbol), then we call the part of the expression that on this occasion appears invariant the function, and the replaceable part its argument.¹² (Frege 1879, § 9)

Although this definition is couched in terms of expressions, in 1891 Frege emphasized the distinction between an expression and its reference and made it clear that functions are what are *denoted* by the corresponding syntactic expressions (Frege 1891). He went on to note that his conception of function extends the previous ones in two ways: by enlarging the collections of signs that could be used to construct a functional expression and by enlarging the domain of possible arguments for functions.¹³ Regarding the first extension, Frege allowed signs such as the equality symbol to occur in functional expressions, thus allowing “ $x^3 + x^2 + x + 1 = 0$ ” to be classified as such. Regarding the second extension, Frege wrote:

Not merely numbers, but objects in general, are now admissible; and here persons must assuredly be counted as objects.¹⁴ (Frege 1891, p. 17)

The syntax of Frege’s logical language allows, moreover, for higher-order functionals, which is to say functions that take elements of a domain of functions as arguments. This feature prompted the following observation in the *Begriffsschrift*:

One sees here particularly clearly that the concept of function in Analysis, which in general I have followed, is far more restricted than that developed here.¹⁵ (Frege 1879, § 10)

Although Frege was influenced by Riemann and developments in the theory of complex functions (Tappenden 2006), there are features of his treatment of the function concept that distinguish it from the contemporary mathematical notion. Nonetheless, in the context of the history we have sketched here, Frege’s dramatic expansion was novel and bold.

¹¹ See § 135 in Dedekind (1888). The relationship between the two notions will be discussed in forthcoming work by Wilfried Sieg and the second author.

¹² “Wenn in einem Ausdrucke, dessen Inhalt nicht beurtheilbar zu sein braucht, ein einfaches oder zusammengesetztes Zeichen an einer oder an mehreren Stellen vorkommt, und wir denken es an allen oder einigen dieser Stellen durch Anderes, überall aber durch Dasselbe ersetzbar, so nennen wir den hierbei unveränderlich erscheinenden Theil des Ausdruckes Function, den ersetzbaren ihr Argument.”

¹³ See Beaney (1997, pp. 137 and 140).

¹⁴ “Es sind nicht mehr bloß Zahlen zuzulassen, sondern Gegenstände überhaupt, wobei ich allerdings auch Personen zu den Gegenständen rechnen muß.”

¹⁵ “Man sich hieran besonders klar, dass der Functionsbegriff der Analysis, dem ich mich im Allgemeinen angeschlossen habe, weit beschränkter ist als der hier entwickelte.”

Today, we are apt to look back at the history and wonder what mathematicians before Dedekind and Frege were missing, that is, why they did not realize the extent to which they could simplify terminology and notation and eliminate conceptual clutter. After all, all they had to do was write $f : A \rightarrow B$ to denote a function f between two arbitrary domains, A and B , and recognize sequences, number-theoretic functions, permutations, and transformations as instances of such. But this seems to us to be the wrong question to ask. A better one is this: Why would anyone want to view these patently different entities as instances of a single concept? In practice, they are described in very different ways: Analytic functions were typically given by piecewise analytic expressions; number-theoretic functions were given by implicit algorithms, such as counting the numbers less than and relatively prime to some number; Galois’ substitutions were given by explicit lists; early geometric transformations were given by geometric constructions. Moreover, the different kinds of objects support very different operations: Functions in analysis can sometimes be composed, but sequences and series cannot; substitutions (and Cayley’s “operations”) can always be inverted, whereas number-theoretic functions generally cannot; one can sum an expression over all the permutations of a finite set, while one certainly cannot sum an expression over all the functions from \mathbb{R} to \mathbb{R} ; notions of continuity and differentiability make no sense outside the realm of analysis, and so on. In a discussion of the function concept in analysis, Renaud Chorlay nicely sums up the state of affairs as follows:

As du Bois-Reymond strikingly put it, the “most general” function, the “arbitrary function,” the “function on which no hypotheses is made” is something about which nothing can be said. It is by no means an object to be studied, it is but an (intensionally) empty place in the whole epistemic configuration: not something to investigate, but a kind of background against which ever more specific function classes can be delineated; meaningless (on the intensional level) because all-encompassing (on the extensional level). (Chorlay n.d.)

This characterization is even more applicable to the notion of a function between two arbitrary domains.

There are invariably start-up costs involved in fashioning coherent means of dealing with new abstract objects and significant concerns. Thus, one should not expect to see an investment in the unification of the function concept until there was either a pressing need or clear benefits to be had. Our study below will suggest that some of these benefits stem from the uniform treatment of algebraic structures composed of functions. For example, the set of automorphisms of a structure form a group under composition, but there is little reason to recognize permutations, geometric transformations, or automorphisms of a field as such until one has particularly useful things to say about groups of automorphisms in general. Similarly, one can view a number of common mathematical constructions as quotienting by the kernel of a suitable function, but there is no point to making the effort to recognize them as such until one realizes that substantial things can be said about the commonality. Alternatively, it is also possible that the unification of the function concept is best viewed as a by-product of the drive toward subsuming mathematical objects under a single unifying foundation, a move which, in turn, might have been encouraged by other mathematical needs.

It is beyond the scope of our study to speculate more than this as to the factors that encouraged the development of a general concept of function and led to its gradual acceptance. But noting the absence of an overarching function concept in the early nineteenth century serves to clarify the nature of the project here. When we consider the history of number-theoretic characters in proofs of Dirichlet's theorem, we are studying the evolving mathematical treatment of *what we now take to be* certain kinds of functions, in the hopes that understanding the forces that guided that evolution will help illuminate the forces that shaped the evolution of modern mathematics more generally. In doing so, we will focus on the way characters were used and the properties that were ascribed to them, while, for the most part, steering clear of the question as to how and why they came to be viewed as instances of the general function concept.

2.2 The evolution of the function concept

We have seen that it was not until the latter half of the nineteenth century that there was any hint of a general notion of a function between any two domains and that it was a while before what we currently take to be instances of the function concept were subsumed under such a general notion. Let us label this trend the *generalization* (or, perhaps, *unification*) of the function concept. The point of this section is to note that this is not the only sense in which the function concept evolved over the course of the century.

To start with, there was also the *extensionalization* of the function concept. Roughly speaking, by an “extensional” view we mean the implicit understanding that a function is completely determined by its extension, that is, the values it takes at all the inputs in its domain. In contrast, by an “intensional” view of the function concept, we mean any way of thinking that presupposes that there is more to a function than its bare extension. For example, let f be the function from real numbers to real numbers given by $f(x) = x^2 + 2x + 1$. Saying that “ f has three terms” or that “the second term of f is $2x$ ” amounts to treating f intensionally, because these statements cannot be couched solely in terms of the values that f takes. Of course, most mathematicians today would acknowledge that such phrases refer to “the expression of f ” or “the definition of f ” or “the representation of f ,” rather than the function f itself. But one of the themes that will emerge below is that this distinction was not clear in the nineteenth century, and fundamental considerations pulled in opposite directions. For concreteness, it is natural to think of functions as expressions, or, at least, entities somehow associated with a concrete representation. But, as we will see, there are also advantages to thinking of operations on functions that depend only on the abstract input–output relation. Our case study will highlight an uneasy tension between these two ways of thinking.

Recall that in Euler's time, a function from the real numbers to the real numbers was expected to have a representation in terms of a certain kind of analytic expression. Arguments involving functions could then make use of such a representation. Precisely this feature of Cauchy's treatment of Fourier series was the target of Dirichlet's criticism in the opening paragraphs of his 1829 paper (Dirichlet 1829) on that same subject. Cauchy, like Dirichlet, was concerned with the question as to when the Fourier series of a real-valued function converges to the value of the original function at a given

point. But Cauchy’s analysis presupposed that the function in question is represented by a power series, which could then be applied to complex arguments. The use of the phrase “arbitrary function” in Dirichlet’s title signaled his intention to lift this restriction. Even though the paper dealt primarily with continuous functions, the hypothesis of continuity was expressed purely extensionally, which is to say in terms of the values that a function takes at its arguments. In other words, Dirichlet took great pains not to treat analytic operations like differentiation as symbolic operations, but, rather, as operations on the functions themselves, viewed extensionally. As Dirichlet famously emphasized in that paper, this renders the concept of function open ended, in that all that is needed is some determinate relationship between input and output values. In particular, this made it possible for him to consider, as a difficult example, functions that take one value on the rationals and another value on the irrational numbers.

This brings us to another aspect of the evolution of the notion of a function, which we will call the *liberalization* of the function concept. Setting aside issues of unification and generality, there is also the issue of the language and methods one helps oneself to in *defining* functions in a particular domain. As Frege later noted,¹⁶ the very act of accepting a definition by cases depending on whether an argument is rational or not was a striking move on Dirichlet’s part. This paves the way to the use of more dramatically nonconstructive methods in describing functions from the reals to the reals, say, in terms of limits and other infinitary operations. Riemann’s contributions to function theory were especially novel in this regard. Some, like Dedekind, hailed the fact that Riemann’s methods made it possible to consider functions without having a particular representation to work with, or method of computation. Others, like Weierstrass, held the Riemannian approach to be defective for just this reason.¹⁷

Finally, there is the issue of the *reification* of the function concept, which is to say the decision to treat functions as objects, on a par with mathematical objects like the natural numbers. In Sect. 4, we will expand on what this amounts to.

We have already indicated that we will have little more to say about the generalization of the function concept. And, since characters are fairly simple combinatorial objects, we can avoid issues having to do with liberal uses of the infinite. We will be quite concerned, however, with issues regarding the reification of characters, and their treatment as extensional objects. Thus, our case study considers some important aspects of the evolution of the function concept, but ignores others, as well as the broader question as to how all the various components are related to one another.

3 Dirichlet’s theorem

Two integers, m and k , are said to be *relatively prime*, or *coprime*, if they have no common factor. In 1837, Dirichlet proved the following:

Theorem 3.1 *If m and k are relatively prime, the arithmetic progression $m, m + k, m + 2k, \dots$ contains infinitely many primes.*

¹⁶ See the discussion in Sect. 7.

¹⁷ See Bottazzini and Gray (2013). For an interesting exploration of the ways that nineteenth-century analysis expanded to incorporate a more liberal understanding of the function concept, see Chorlay (n.d.).

In other words, if m and k are relatively prime, there are infinitely many primes congruent to m modulo k . Dirichlet pointed out that Legendre assumed this fact, without proof, in 1788 (Legendre 1788), when proving the law of quadratic reciprocity. In his *Disquisitiones Arithmeticae*, Gauss noted this gap in Legendre's work after presenting his own proof of the law of quadratic reciprocity, one which does not rely on Theorem 3.1.¹⁸ But Gauss himself was never able to prove that theorem. Dirichlet's own proof is striking, not only due to the fact that it finally established Legendre's conjecture, but also due to its sophisticated use of the methods of analysis in establishing a purely number-theoretic assertion. Dirichlet noted (Dirichlet 1837a, pp. 309–310) that his method was inspired by a proof that there are infinitely many primes due to (Euler 1748, Chapter XV), though Dirichlet's ideas go considerably beyond Euler's. In this section, we will describe Euler's proof and then define the modern notion of a group-theoretic character, which supports the generalization to arbitrary characters. We will then describe contemporary proofs of Dirichlet's theorem, using modern terminology and notation, as presented in textbooks such as Everest and Ward (2005, pp. 207–224). In Sect. 4, we will use this presentation to frame our discussion of the historical development, which occurs in Sects. 5 and 6.

There is a sense, however, in which our modern presentation is old-fashioned. The use of m and k for the initial value and common difference in the statement of Theorem 3.1 is due to Dirichlet and was picked up by a number of successive authors, including Dedekind and Hadamard. Modern presentations are more apt to use a and d , but although the use of m and k may feel alien to readers familiar with contemporary textbook proofs, it will facilitate the historical comparisons later on. (Our reason for using the variable q to range over the prime numbers will similarly become clear when we discuss Dirichlet's original proof.)

3.1 Euler's proof that there are infinitely many primes

In the *Elements*, Euclid proved that there are infinitely many primes, but his proof does not provide much information about how they are distributed. Euler, in his *Introductio in Analysin Infinitorum* (Euler 1748), proved the following:

Theorem 3.2 *The series $\sum_q \frac{1}{q}$ diverges, where the sum is over all primes q .*

This implies that there are infinitely many primes, but also says something more about their density. For example, since we know that the series $\sum_n \frac{1}{n^2}$ is convergent, it tells us that, in a sense, there are “more” primes than there are squares.

Euler's proof of Theorem 3.2 centers around his famous zeta function,

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s},$$

defined for a real variable s . (The zeta function was later extended by Riemann to the entire complex plane via analytic continuation.) It is not hard to show that the

¹⁸ See Gauss' remarks in Friedrich Gauss (1801, § 150).

series $\zeta(s)$ converges uniformly on the interval $[a, \infty)$, where a is any number strictly greater than 1. For $s > 1$, the infinite sum can also be expressed as an infinite product:

$$\sum_{n=1}^{\infty} n^{-s} = \prod_q \left(1 - \frac{1}{q^s}\right)^{-1},$$

where the product is over all primes q . This is known as the *Euler product formula*. Roughly, this holds because we can write each term of the product as the sum of a geometric series,

$$\left(1 - \frac{1}{q^s}\right)^{-1} = 1 + q^{-s} + q^{-2s} + \dots$$

and then expand the product into a sum. The unique factorization theorem tells us that every integer $n > 1$ can be written as a product $q_1^{i_1} \cdot q_2^{i_2} \dots q_k^{i_k}$, and so the term $n^{-s} = q_1^{-i_1 s} \cdot q_2^{-i_2 s} \dots q_k^{-i_k s}$ will occur exactly once in the expansion. Since we are dealing with infinite sums and products, the Euler product formula implicitly makes a statement about limits, and some care is necessary to make the argument precise, but this is not hard to do.

If we take the logarithm of each side of the product formula and appeal to properties of the logarithm function, we obtain

$$\log \sum_{n=1}^{\infty} n^{-s} = \sum_q -\log \left(1 - \frac{1}{q^s}\right).$$

Using the Taylor series expansion

$$\log(1 - x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

and changing the order of summations yields

$$\log \sum_{n=1}^{\infty} n^{-s} = \sum_q \frac{1}{q^s} + \sum_{n=2}^{\infty} \frac{1}{n} \sum_q \frac{1}{q^{ns}}.$$

At this stage, keep in mind that we want to show that $\sum_q \frac{1}{q}$ diverges, and notice that the first term on the right-hand side of the above equation is $\sum_q \frac{1}{q^s}$. Thus, we should consider what happens as s tends to 1 from above. It is not hard to show that the second term on the right-hand side is bounded by a constant that is independent of s , a fact that can be expressed using “big O” notation as follows:

$$\log \sum_{n=1}^{\infty} n^{-s} = \sum_q \frac{1}{q^s} + O(1). \quad (1)$$

As s approaches 1 from above, the left-hand side clearly tends to infinity. Thus, the right-hand side, $\sum_q \frac{1}{q^s}$, must also tend to infinity, which implies that $\sum_q \frac{1}{q}$ diverges.

3.2 Group characters

Just as Euler's proof shows that there are infinitely many primes by establishing that the series $\sum_q \frac{1}{q}$ is divergent, proofs of Dirichlet's theorem establish that there are infinitely many primes q congruent to m modulo k by showing that the series $\sum_{q \equiv m \pmod{k}} \frac{1}{q}$ is divergent. Remember that we are assuming that m and k are relatively prime. In general, the residues modulo k that are relatively prime to k form a multiplicative group. To adapt the argument above, we need series that are more refined than the zeta function, and a device that enables us to focus attention on the integers that are congruent to m modulo k . This is where the notion of a group-theoretic character comes in.

Notice that the expression $\sum_{q \equiv m \pmod{k}} \frac{1}{q}$ can be written $\sum_q \frac{f(q)}{q}$, where $f(n)$ is equal to 1 if n is congruent to m modulo k and 0 otherwise. With this observation, it is natural to try to emulate Euler's argument with $\sum_{n=1}^{\infty} f(n)n^{-s}$ in place of the ζ function, $\sum_{n=1}^{\infty} n^{-s}$. We will see below that an analogue of the Euler product formula holds if f is *completely multiplicative*, which is to say it satisfies $f(uv) = f(u)f(v)$ for every u and v . The problem is that the particular f just described does not have this property. But Corollary 3.5 below shows that we can decompose f into a sum of functions that do have this property and apply the Euler product formula to each component.

Let G be a finite abelian group, written multiplicatively, with an identity denoted by 1. A *character* χ on G is a homomorphism from G to the nonzero complex numbers, \mathbb{C}^* . In other words, a character χ is a nonzero function satisfying $\chi(g_1 g_2) = \chi(g_1)\chi(g_2)$ for every g_1 and g_2 in G . Since $\chi(1) = \chi(1 \cdot 1) = \chi(1)\chi(1)$ and $\chi(1)$ is nonzero, we have $\chi(1) = 1$. Since G is a finite group, for every g there is a least $n > 1$ satisfying $g^n = 1$; this n is called the *order* of g and denoted $o(g)$. The fact that $\chi(g)^{o(g)} = \chi(g^{o(g)}) = \chi(1) = 1$ means that $\chi(g)$ is a "root of unity" for every g in G . The character that is equal to 1 for every g in G is called the *trivial character* and denoted by χ_0 .

Define the product $\chi \cdot \psi$ of two characters pointwise, by

$$(\chi \cdot \psi)(g) = \chi(g)\psi(g),$$

for every g in G . This multiplication is commutative, and we have $\chi \cdot \chi_0 = \chi$ for every character χ , which is to say χ_0 is a multiplicative identity. Recall that if ω is any complex root of unity, its complex conjugate, $\bar{\omega}$, is also a root of unity, satisfying $\omega\bar{\omega} = 1$. We can also lift the operation of conjugation to characters, defining $\bar{\chi}$ by the equation $\bar{\chi}(g) = \overline{\chi(g)}$ for each g . Then clearly we have $\chi \cdot \bar{\chi} = \chi_0$. In other words, the set of characters on G forms an abelian group, with $\chi^{-1} = \bar{\chi}$. We will denote this group \hat{G} .

The following theorem is fundamental:

Theorem 3.3 *If G is any finite abelian group, \widehat{G} is isomorphic to G . In particular, $|\widehat{G}| = |G|$.*

To see this, first consider the case where $G = \langle g \rangle = \{1, g, g^2, \dots, g^{n-1}\}$ is a cyclic group generated by an element g of order n . Then any character χ on G has to map g to an n th root of unity, ω . This determines the behavior of χ completely, since then we have $\chi(g^i) = \omega^i$ for every i . Thus, we can let χ_ω denote the unique character that maps g to ω . But now if we let $\omega = e^{2\pi i/n}$, then ω is what is known as a “primitive root of unity,” which is to say that all the roots of unity are given by $1, \omega, \omega^2, \dots, \omega^{n-1}$. Notice that these roots form a multiplicative group that is isomorphic to G . It is easy to verify that the map that sends the element g^i of G to the character χ_{ω^i} of \widehat{G} is an isomorphism.

In the more general case, we appeal to the structure theorem for finite abelian groups, which says that any such group G can be written as a product $G_1 \times \dots \times G_k$ of cyclic groups. This means that every element g of G can be written uniquely as a product $g = g_1 g_2 \dots g_k$, where each g_i is in G_i . Given characters $\chi_1, \chi_2, \dots, \chi_k$ on G_1, G_2, \dots, G_k , respectively, one gets a character χ on G defined by $\chi(g) = \chi_1(g_1)\chi_2(g_2)\dots\chi_k(g_k)$, where $g = g_1 g_2 \dots g_k$ is the decomposition described above. Moreover, it is not hard to show that every character on G arises in this way. This shows that \widehat{G} is isomorphic to $\widehat{G}_1 \times \widehat{G}_2 \times \dots \times \widehat{G}_k$. By the analysis of the cyclic case, the latter is, in turn, isomorphic to $G_1 \times G_2 \times \dots \times G_k$, and hence to G .

The following theorem expresses two important properties, known as the “orthogonality relations” for group characters.

Theorem 3.4 *Let G be finite abelian group. Then for any character χ in \widehat{G} , we have*

$$\sum_{g \in G} \chi(g) = \begin{cases} |G| & \text{if } \chi = \chi_0 \\ 0 & \text{if } \chi \neq \chi_0. \end{cases}$$

and for any element g of G , we have

$$\sum_{\chi \in \widehat{G}} \chi(g) = \begin{cases} |G| & \text{if } g = 1_G \\ 0 & \text{if } g \neq 1_G. \end{cases}$$

The first equation clearly holds when χ is the trivial character, χ_0 , since, in this case, each term of the sum is equal to 1. Otherwise, pick h such that $\chi(h) \neq 1$ and note

$$\chi(h) \sum_{g \in G} \chi(g) = \sum_{g \in G} \chi(hg) = \sum_{g \in G} \chi(g),$$

since hg ranges over the elements of G as g does. Since $\chi(h) \neq 1$, we must have $\sum_{g \in G} \chi(g) = 0$. The second equation can be established in a similar way.

The orthogonality relations make it possible to do “finite Fourier analysis,” in the following sense: If f is any function from G to the complex numbers and we define the “Fourier transform” \hat{f} of f by $\hat{f}(\chi) = \sum_g f(g)\chi(g)$, then f can be recovered from

its Fourier transform: $f = \frac{1}{|G|} \sum_{\chi} \hat{f}(\chi) \chi$. This shows, in particular, that any function from G to the complex numbers can be written as a linear combination of characters. The second orthogonality relation also provides the following useful corollary:

Corollary 3.5 *For any $g, h \in G$, we have the following:*

$$\sum_{\chi \in \hat{G}} \chi(g) \overline{\chi(h)} = \begin{cases} |G| & \text{if } g = h \\ 0 & \text{if } g \neq h \end{cases}$$

This follows from the fact that we have

$$\sum_{\chi \in \hat{G}} \chi(g) \overline{\chi(h)} = \sum_{\chi \in \hat{G}} \chi(g) \chi(h)^{-1} = \sum_{\chi \in \hat{G}} \chi(gh^{-1}) = \begin{cases} |G| & \text{if } g = h \\ 0 & \text{if } g \neq h \end{cases}.$$

This corollary will enable us to focus on the residue class of m modulo k in the proof of Dirichlet's theorem.

3.3 Dirichlet characters and L -series

Let k be an integer greater than or equal to 1. It is a fundamental theorem of number theory that an integer n is relatively prime to k if and only if n has a multiplicative inverse modulo k , in other words if and only if there is some n' such that $nn' \equiv 1 \pmod{k}$. This implies that the residue classes of integers modulo k that are relatively prime to k form a group, denoted $(\mathbb{Z}/k\mathbb{Z})^*$, with multiplication modulo k . The cardinality of $(\mathbb{Z}/k\mathbb{Z})^*$, that is, the number of residues relatively prime to k , is denoted by $\varphi(k)$, and the function φ is called the *Euler phi function*.

Any character χ on $(\mathbb{Z}/k\mathbb{Z})^*$ can be “lifted” to a function X from \mathbb{Z} to \mathbb{C} defined by

$$X(n) = \begin{cases} \chi(n \bmod k) & \text{if } n \text{ is relatively prime to } k \\ 0 & \text{otherwise.} \end{cases}$$

Such a function is called a *Dirichlet character modulo k* . Dirichlet characters are completely multiplicative, which is to say $X(mn) = X(m)X(n)$ for every m and n in \mathbb{Z} . Mathematicians typically use the symbol χ to range over Dirichlet characters, blurring the distinction between such functions and their group-character counterparts. This is harmless, since there is a one-to-one correspondence between the two, and so we will adopt this practice as well.

We can now generalize the method of Euler's proof. Roughly speaking, we need a variant of the zeta function that will allow us to focus on primes in a particular residue class modulo k . To that end, given a Dirichlet character χ modulo k , define the *Dirichlet L -function* or *L -series*,

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s},$$

where s is any complex number. This formal series will converge whenever $\Re(s) > 1$, that is, the real part of s is greater than one. And just as $\sum_{n=1}^{\infty} \frac{1}{q^s}$ can be written as a product via the Euler product formula, each $L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$ has a useful product expansion.

Theorem 3.6 *Let χ be a Dirichlet character modulo k . Then the L -function associated with χ has an Euler product expansion for $\Re(s) > 1$,*

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_q \left(1 - \frac{\chi(q)}{q^s}\right)^{-1} = \prod_{q \nmid k} \left(1 - \frac{\chi(q)}{q^s}\right)^{-1}.$$

The last identity follows from the fact that, in the product, we can ignore those primes q that divide k , since $\chi(q) = 0$ for such q . With the product formula in place, we can sketch a proof of Dirichlet’s theorem.

3.4 Contemporary proofs of Dirichlet’s theorem

Recall that we want to prove that there are infinitely many primes q such that $q \equiv m \pmod{k}$, where m and k are relatively prime. As in the proof that $\sum_q \frac{1}{q^s}$ diverges, we begin by taking logarithms of both sides of the Euler product expansion for $L(s, \chi)$, where χ is a Dirichlet character modulo k :

$$\log L(s, \chi) = - \sum_{q \nmid k} \log \left(1 - \frac{\chi(q)}{q^s}\right).$$

As before, we make use of the Taylor series expansion for the logarithm on the right-hand side of the above equation to obtain:

$$\begin{aligned} \log L(s, \chi) &= \sum_{q \nmid k} \sum_{j=1}^{\infty} \frac{1}{j} \frac{\chi(q^j)}{q^{sj}} \\ &= \sum_{q \nmid k} \frac{\chi(q)}{q^s} + \sum_{q \nmid k, j=2}^{\infty} \frac{1}{j} \frac{\chi(q^j)}{q^{sj}}. \end{aligned}$$

One can show that the second term in the expression is bounded by a constant that is independent of s and χ , which can be expressed as follows:

$$\log L(s, \chi) = \sum_{q \nmid k} \frac{\chi(q)}{q^s} + O(1).$$

Now comes the crucial use of Corollary 3.5 to pick out the primes in the relevant residue class. We multiply each side of the above equation by $\chi(m)$ and then take the sum of these over all the Dirichlet characters modulo k . (Recall that we can identify each

Dirichlet character with the corresponding group character, that is, the corresponding element of $(\widehat{\mathbb{Z}/k\mathbb{Z}})^*$.) Thus, we have:

$$\sum_{\chi \in (\widehat{\mathbb{Z}/k\mathbb{Z}})^*} \overline{\chi(m)} \log L(s, \chi) = \sum_{\chi \in (\widehat{\mathbb{Z}/k\mathbb{Z}})^*} \overline{\chi(m)} \sum_{q \nmid k} \frac{\chi(q)}{q^s} + O(1).$$

To simplify this expression, we exchange the summations on the right-hand side and appeal to Corollary 3.5. Since the cardinality of the group $(\mathbb{Z}/k\mathbb{Z})^*$ is $\varphi(k)$, we obtain

$$\sum_{\chi \in (\widehat{\mathbb{Z}/k\mathbb{Z}})^*} \overline{\chi(m)} \log L(s, \chi) = \varphi(k) \sum_{q \equiv m \pmod{k}} \frac{1}{q^s} + O(1). \quad (2)$$

This is analogous to Eq. (1) in Euler's proof. Our goal is once again to show that the left-hand side tends to infinity as s approaches 1 from above; this implies that the right-hand side tends to infinity, which, in turn, implies that there are infinitely many primes q that are congruent to m modulo k . However, now the left-hand side is considerably more complicated than the expression $\log \sum_{n=1}^{\infty} n^{-s}$ in Euler's proof.

To show that $\sum_{\chi \in (\widehat{\mathbb{Z}/k\mathbb{Z}})^*} \overline{\chi(m)} \log L(s, \chi)$ tends to infinity as s approaches 1, we divide the characters into three classes, as follows:

1. The first class contains only the principal character χ_0 , which takes the value of 1 for all arguments that are relatively prime to k , and 0 otherwise.
2. The second class consists of all those characters that take only real values (i.e., 0 or ± 1), other than the principal character.
3. The third class consists of those characters that take at least one complex value.

It is not difficult to show that $L(s, \chi_0)$ has a simple pole at $s = 1$, which implies that the term $\overline{\chi_0(m)} \log L(s, \chi_0)$ approaches infinity as s approaches 1. The real work involves showing that for all the other characters χ , $L(s, \chi)$ has a finite nonzero limit. This implies that the other terms in the sum approach a finite limit, and so the entire sum approaches infinity.

For characters in the third class, that is, the characters that take on at least one complex value, the result is not difficult. For characters in the second class, the result is much harder, and Dirichlet used deep techniques from the theory of quadratic forms to obtain it. In the years that followed, other mathematicians found alternative, and simpler, ways of handling this case. But even in modern presentations, this case remains the most substantial and technically involved part of the proof.

Our presentation has been thoroughly "modern." In the next section, we will consider some of the methodological features of the proof that make it so. This will enable us to draw interesting contrasts with Dirichlet's proof and then explore the way that presentations of Dirichlet's theorem gradually took on such a modern character.

4 Modern aspects of contemporary proofs

Although our presentation uses contemporary terminology and notation, there is a sense in which it is a faithful description of Dirichlet’s 1837 proof. Dirichlet did not, and could not, rely on a general notion of group character, as the general notion of a group was not articulated before Cayley did so in 1854 (Cayley 1854), and was not brought into general currency until Kronecker’s 1870 paper (Kronecker 1870), which first presented the structure theorem for finite abelian groups.¹⁹ Another important difference is that even though Dirichlet’s argument used complex numbers in a central way, there was less established background in complex analysis than is available today, and Dirichlet tended to reduce the calculations to real analysis whenever possible. Thus, his variable s ranged over real numbers, and his calculations involve real-valued sines and cosines where today we are comfortable sticking with the complex exponential. In addition, we have already noted that many of the technical details were streamlined over the years. Despite all this, the outline above characterizes the central ideas of his proof, and most mathematicians would not find it unreasonable to say that is, essentially, how he obtained the result.

But, as we will see in Sect. 5, there is one very striking difference: In Dirichlet’s original presentations, there is no mention of characters at all. That is, Dirichlet’s papers contain certain expressions that we now recognize as values of the various characters, and summations that are *tantamount* to summing over all the characters. But, the characters themselves are only objects that we project back into the argument from our current understanding. They hover over the page as shade-like premonitions, ghosts of mathematics yet to come.

In Sect. 6, we will consider presentations of Dirichlet’s theorem given by Dedekind, de la Vallée-Poussin, Hadamard, Kronecker, and Landau and see how the characters were gradually brought to life. We will see that many of the benefits of giving characters a substantive embodiment are notational and pragmatic, but that is not to say that they are *merely* notational and pragmatic: Treating characters as bona fide objects comes with serious mathematical constraints and obligations and provides conceptual reorientations that have great bearing on the kind of mathematics we do and the way we do it. We will argue that the reification of the notion of a character is a prototypical instance of the conceptual changes that are hallmarks of the transition to “modern” mathematical thought, and that understanding how and why the changes came about shed light on the way we do mathematics today.

But the observations above present us with a terminological conundrum: Should we describe the various historical texts as “versions” or “presentations” of Dirichlet’s proof, or different proofs entirely? Having raised this issue, we will, for the most part, set it aside and be fairly cavalier with our terminology. Since our specific concern is to study the way that language, conceptualization, and inferential practice evolved over the years, and the effects that had on the mathematics, we need not explicitly address the question as to when it is proper to consider two proofs essentially the same or essentially different.

¹⁹ See Wussing (1984) for the history of group theory.

Talking about historical texts in modern terms is difficult, and it is always misleading to portray the history of mathematics as a muddled and inefficient attempt to arrive at the contemporary enlightened view. We hope we have not fallen into this trap. If there is anything that deserves to be treated as a rational pursuit, mathematics should count as such, and so one would expect there to be good reasons that we do mathematics the way we do. At the same time, there are often good reasons to question the way we do mathematics today and recognize there are trade-offs involved in the historical decisions that were made. Comparing mathematical texts from different historical eras and trying to understand what has changed provides a fruitful way of understanding the values that drive mathematical change. But it is often easier to draw contrasts by starting with the mathematics with which we are most familiar, and so having discussed contemporary approaches to Dirichlet's theorem, let us foreshadow some of the contrasts we wish to consider.

In Sect. 2, we noted that characters (whether we refer to group characters or Dirichlet characters) are instances of the contemporary function concept, a concept that evolved significantly over the course of the nineteenth century. These are some of the salient features of the treatment of characters in our modern presentation:

1. Group characters are given an abstract, axiomatic definition as functions that satisfy the homomorphism property, and the Dirichlet characters are introduced as a natural extension of the notion.
2. In particular, one defines the set of characters modulo k extensionally. Only later does one show that this set is finite and provide explicit ways of describing and enumerating them.
3. Characters are studied in their own right, and their general properties are enunciated in propositions and theorems.
4. One sums over sets of characters, without needing representations for any particular one. More generally, one characterizes operations on characters (such as the product of two characters) extensionally, and not in terms of their representations.
5. Characters appear as arguments to other functions, namely the L -functions. In particular, it is clear from the definition that $L(s, \chi)$ depends only on the extension of χ .
6. One defines various sets of characters extensionally, for example, distinguishing the trivial, real, and complex characters in terms of the values they assume. More generally, one typically carries out arguments without making reference to any particular representation.
7. The characters modulo k are viewed as elements of an algebraic structure, namely a group, with multiplication defined pointwise and the trivial character serving as identity.

In Sect. 5, we will discuss Dirichlet's original proof, and in Sect. 6, we will consider the way the proof was gradually transformed to reflect our contemporary understanding. We will see that proofs along the way possess various subsets of the properties just enumerated and that in some cases, the authors are noticeably squeamish, or at least self-conscious, of these features. Importantly, Dirichlet's original proof has *none* of the properties just enumerated, providing a clear contrast to the style of presentation that is common today.

One way of characterizing the difference between Dirichlet’s original proof and our modern presentation is to say that, in the latter, characters are treated as full-fledged mathematical objects, whereas there are no such objects in Dirichlet’s version. Elsewhere (Avigad and Morris n.d.), we reflect in greater detail on what it means to say that a piece of mathematics sanctions certain entities as “objects.” Here, let us merely summarize some of the senses in which this can be said to be the case in the modern proofs of Dirichlet’s theorem:

1. Characters fall under a recognized grammatical category, which allows us to state things about them and define operations and predicates on them.
2. There is a clear understanding of what it means for two expressions to represent the *same* character, namely that they take the same values at all arguments. Linguistic conventions ensure that expressions occurring in a proof respect this “sameness.” For example, the expression $L(s, \chi)$ does not depend on the data used to represent χ .
3. One can quantify and sum over characters; in logical terms, they can fall under the range of a bound variable.
4. One can define functions that take characters as arguments and sets of characters; indeed, characters can be elements of arbitrary algebraic structures.

What these features have in common is that they are fundamental to the way we do mathematics, bearing upon the proper use of mathematical language and inference at a very low level.

5 Dirichlet’s original proof

We have already asserted that characters appear only “implicitly” in Dirichlet’s original proof (Dirichlet 1837b). There is nothing mysterious about this: What we mean is that, in Dirichlet’s proof, there are certain symbolic expressions that we now recognize as denoting the values of characters and that, moreover, some of Dirichlet’s calculations and inferences invoke what we now recognize as general properties of characters.

Let us spell out the details. Like Dirichlet, we will first consider the case where the common difference in the arithmetic progression is a prime number, denoted by p instead of k . It was well known in the nineteenth century that one can always find a *primitive element* modulo p , which is to say an element c , such that the $p - 1$ residues $c^0, c^1, c^2, \dots, c^{p-2}$ modulo p yield all the nonzero residues, $1, 2, \dots, p - 1$ (not necessarily in the same order). In modern terms, we would say that the group $(\mathbb{Z}/p\mathbb{Z})^*$ of units modulo p is cyclic, generated by the residue class of c . In more elementary terms, this amounts to saying that for every integer n , there is a number γ_n with the property that $c^{\gamma_n} \equiv n \pmod{p}$. We saw in Sect. 3.2 that if χ is a Dirichlet character modulo p (i.e., χ corresponds to character on $(\mathbb{Z}/p\mathbb{Z})^*$), then $\chi(c)$ is a p th root of unity, say, ω ; moreover, χ is entirely determined by ω , in the sense that for every n relatively prime to p , $\chi(n) = \omega^{\gamma_n}$. Thus, Dirichlet simply wrote ω^{γ_n} where we would write $\chi(n)$. The notation presupposes that one has fixed a choice of the primitive element, c , though any primitive element will work equally well.

So far, so good. In the more general case where the modulus is a composite number k , however, things get more complicated. First, write k as a product of primes,

$$k = 2^\lambda p_1^{\pi_1} p_2^{\pi_2} \dots p_j^{\pi_j}$$

where each p_i is an odd prime and π_i is greater than or equal to 1. Then the group of units modulo k is isomorphic to the product of the groups of units modulo each term in the factor. Gauss had already shown if p is an odd prime and π is an integer greater than or equal to 1, then one can more generally find a primitive element c modulo p^π . This means that the residue class of c generates the cyclic group $(\mathbb{Z}/p^\pi \mathbb{Z})^*$, or, equivalently, for every n relatively prime to p , there is a γ_n such that $c^{\gamma_n} \equiv n \pmod{p^\pi}$. Thus, we can choose primitive elements c_1, \dots, c_j corresponding to $p_1^{\pi_1}, p_2^{\pi_2}, \dots, p_j^{\pi_j}$. If $\lambda \geq 3$, however, there is no primitive element modulo 2^λ . Rather, $(\mathbb{Z}/2^\lambda)^*$ is a product of two cyclic groups, and for every n relatively prime to 2^λ , there are an α_n and β_n such that $(-1)^{\alpha_n} 5^{\beta_n} \equiv n \pmod{2^\lambda}$. Thus, for any n relatively prime to k , we can write

$$n \equiv (-1)^{\alpha_n} 5^{\beta_n} c_1^{\gamma_{1,n}} c_2^{\gamma_{2,n}} \dots c_j^{\gamma_{j,n}} \pmod{k}$$

where each $\gamma_{i,n}$ is the index n relative to $p_i^{\pi_i}$. As above, if we choose appropriate roots of unity $\theta, \varphi, \omega_1, \omega_2, \dots, \omega_j$, we obtain a character

$$\chi(n) = \theta^{\alpha_n} \varphi^{\beta_n} \omega_1^{\gamma_{1,n}} \omega_2^{\gamma_{2,n}} \dots \omega_j^{\gamma_{j,n}}. \quad (3)$$

And, once again, every character is obtained in this way. We should note that Dirichlet used the notation p, p', \dots rather than p_1, \dots, p_j to denote the sequence of odd primes. Moreover, he used the notation $\alpha, \beta, \gamma, \gamma', \dots$ to denote the indices, suppressing the dependence on n . Thus, Dirichlet wrote $\theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots$ for the expression we have denoted $\chi(n)$ above, leaving it up to us to keep in mind that α, β, \dots depend on n .

To summarize, in the simple case of a prime modulus p , Dirichlet fixed a primitive element modulo c and represented each character χ in terms of a p th root of unity, ω . In that case, the value $\chi(n)$ is given by ω^{γ_n} . In the more general case of a composite modulus k , Dirichlet fixed primitive elements modulo the terms of the prime factorization of k and represented each character χ in terms of a sequence $\theta, \varphi, \pi, \pi'$ of roots of unity. In that case, the value $\chi(n)$ was written $\theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots$, suppressing the information that the exponents $\alpha, \beta, \gamma, \gamma', \dots$ depend on n .

Recall that our contemporary presentation had little to say about *particular* characters, other than the trivial character, χ_0 . Rather, characters appear as arguments to the L -functions, $L(s, \chi)$, and the proof considers summations over the set of all characters. Let us now consider how Dirichlet handled these as well.

Again, with Dirichlet, we begin with the easier case where the common difference in the arithmetic progression is a prime, p . Recall that, in that case, each character χ corresponds to a p th root of unity, ω . Dirichlet stated the Euler product formula as follows:

We therefore have the equation

$$\prod \frac{1}{1 - \omega^\gamma \frac{1}{q^s}} = \sum \omega^\gamma \frac{1}{n^s} = L, \quad (4)$$

where the multiplication sign ranges over the whole series of primes with the sole exception of p , while the summation involves all the integers from 1 to ∞ that are not divisible by p . The letter γ denotes γ_p on the left, and γ_n on the right.²⁰ (Dirichlet 1837b, p. 3)

Compare this to the statement of Theorem 3.6 above. Since there are $p - 1$ distinct $p - 1$ st roots of unity, Dirichlet continued:

The equation just found represents $p - 1$ different equations, which are obtained by replacing ω with its $p - 1$ values. It is known that these $p - 1$ different values can be represented as powers of one such Ω , chosen appropriately, so that the values are then:

$$\Omega^0, \Omega^1, \Omega^2, \dots, \Omega^{p-2}.$$

In accordance with this representation, we will write the different values L of the series or product as:

$$L_0, L_1, L_2, \dots, L_{p-2},$$

...²¹ (Dirichlet 1837b, p. 3)

²⁰ "Man hat daher die Gleichung:

$$\prod \frac{1}{1 - \omega^\gamma \frac{1}{q^s}} = \sum \omega^\gamma \frac{1}{n^s} = L. \quad (4)$$

wo sich die Multiplicationszeichen auf die ganze Reihe der Primzahlen, mit alleiniger Ausnahme von p , erstreckt, während die Summation sich auf alle ganzen Zahlen von 1 bis ∞ bezieht, welche nicht durch p teilbar sind. Der Buchstabe γ bedeutet auf der ersten Seite γ_q , auf der zweiten dagegen γ_n ." (Dirichlet 1837b, pp. 317–318) We have replaced Dirichlet's equation number with our own, and throughout this section we have modified the translation cited in Dirichlet (1837b).

²¹ "Die eben gefundene Gleichung repräsentirt $p - 1$ verschiedene Gleichungen, welche man erhält, wenn man für ω seine $p - 1$ Werthe setzt. Bekanntlich lassen sich diese $p - 1$ verschiedenen Werthe durch die Potenzen von einem derselben Ω darstellen, wenn dieser gehörig gewählt wird, und sind dann:

$$\Omega^0, \Omega^1, \Omega^2, \dots, \Omega^{p-2}.$$

Wir werden, dieser Darstellung entsprechend, die verschiedenen Werthe L der Reihe oder des Productes mit:

$$L_0, L_1, L_2, \dots, L_{p-2}$$

bezeichnen...." (Dirichlet 1837b, p. 318)

Notice that Dirichlet says that the Euler product formula “represents $p - 1$ different equations,” rather than thinking of it as a single equation parametrized by ω .

In the more general case where the common difference is some composite k , Dirichlet’s procedure is completely analogous. First, he demonstrated that the Euler product formula holds:

$$\prod \frac{1}{1 - \theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots \frac{1}{q^s}} = \sum \theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots \frac{1}{n^s} = L, \quad (5)$$

where the multiplication sign ranges over all primes, with the exclusion of $2, p, p', \dots$, and the summation ranges over all the positive integers that are not divisible by any of the primes $2, p, p', \dots$. The system of indices $\alpha, \beta, \gamma, \gamma', \dots$ on the left side corresponds to the number q , and on the right side to the number n . The general Eq. (5), in which the different roots $\theta, \varphi, \omega, \omega', \dots$ can be combined with one another arbitrarily, clearly contains K -many particular equations.²² (Dirichlet 1837b, p. 17)

Note, again, Dirichlet’s characterization of the general equation as “containing” the particular instances. Here, K is what we have called $\varphi(k)$, the cardinality of the group $(\mathbb{Z}/k\mathbb{Z})^*$. Dirichlet went on to note that we can choose primitive roots of unity $\Theta, \Phi, \Omega, \Omega', \dots$ so that all choices of $\theta, \varphi, \omega, \omega', \dots$ can be expressed as powers of these,

$$\theta = \Theta^a, \varphi = \Phi^b, \omega = \Omega^c, \omega' = \Omega'^c, \dots,$$

just as in the simpler case. He wrote that we can thus refer to the L -series in a “convenient” (“bequem”) way, as $L_{a,b,c,c',\dots}$, where a, b, c, c', \dots are the exponents of the chosen primitive roots. Notice that the representations just described depend on fixed, but arbitrary, choices of the primitive roots of unity, as well as fixed but arbitrary generators of the cyclic groups. Modulo those choices, we have parameters a, b, c, c', \dots that vary to give us all the characters, and for each choice of a, b, c, c', \dots , we have an explicit expression that tells us the value of the character at n . For Dirichlet, summing over characters therefore amounted to summing over all possible choices of this representing data.

²² “

$$\prod \frac{1}{1 - \theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots \frac{1}{q^s}} = \sum \theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots \frac{1}{n^s} = L, \quad (5)$$

wo sich das Multiplikationszeichen auf die ganze Reihe der Primzahlen, mit Ausschluss von $2, p, p', \dots$, und das Summenzeichen auf alle positiven ganzen Zahlen, welche durch keine der Primzahlen $2, p, p', \dots$ theilbar sind, erstreckt. Das System der Indices $\alpha, \beta, \gamma, \gamma', \dots$ entspricht auf der ersten Seite der Zahl q , auf der zweiten Seite der Zahl n . Die allgemeine Gleichung (5), in welcher die verschiedenen Wurzeln $\theta, \varphi, \omega, \omega', \dots$ auf irgend eine Weise mit einander combinirt werden können, enthält offenbar eine Anzahl K besonderer Gleichungen.” (Dirichlet 1837b, pp. 336–337) We have replaced Dirichlet’s equation number with our own.

In the special case where the common difference is a prime, p , Dirichlet ran through calculations similar to those described in Sect. 3.4 to obtain the following identity:

$$\begin{aligned} & \sum \frac{1}{q^{1+\rho}} + \frac{1}{2} \sum \frac{1}{q^{2+2\rho}} + \frac{1}{3} \sum \frac{1}{q^{3+3\rho}} + \cdots \\ &= \frac{1}{p-1} (\log L_0 + \Omega^{-\gamma_m} \log L_1 + \Omega^{-2\gamma_m} \log L_2 + \cdots + \Omega^{-(p-1)\gamma_m} \log L_{p-2}). \end{aligned}$$

He has essentially arrived at Eq. (2) in Sect. 3.4, which reads as follows:

$$\sum_{\chi \in (\mathbb{Z}/k\mathbb{Z})^*} \overline{\chi(m)} \log L(s, \chi) = \varphi(k) \sum_{q \equiv m \pmod{k}} \frac{1}{q^s} + O(1) \quad (2)$$

In the case at hand, k is p , in which case $\varphi(k)$, the cardinality of $(\mathbb{Z}/k\mathbb{Z})^*$, is equal to $p-1$. To facilitate the comparison, switch the right-hand side of (2) with the left-hand side (convention dictates that the term $O(1)$ stays on the right side), divide through by $\varphi(k)$, and note all of the following. First, Dirichlet used $1+\rho$ in place of s to denote the quantity that approaches 1 from above. Second, the sum

$$\frac{1}{2} \sum \frac{1}{q^{2+2\rho}} + \frac{1}{3} \sum \frac{1}{q^{3+3\rho}} + \cdots$$

has been absorbed into the term $O(1)$ in Eq. (2); indeed, Dirichlet’s next move was to note that this sum is bounded by a constant. Finally, if $\chi_i(m)$ is equal to $\Omega^{i\gamma_m}$, then the complex conjugate $\overline{\chi_i(m)}$ is equal to $\Omega^{-i\gamma_m}$, so the expression $\Omega^{-i\gamma_m} \log L_i$ would be expressed in our notation as $\overline{\chi_i(m)} L(s, \chi_i)$.

When Dirichlet considered the more general case, he arrived at the analogous result:

$$\begin{aligned} & \sum \frac{1}{q^{1+\rho}} + \frac{1}{2} \sum \frac{1}{q^{2+2\rho}} + \frac{1}{3} \sum \frac{1}{q^{3+3\rho}} + \cdots \\ &= \frac{1}{K} \sum \Theta^{-\alpha_m a} \Phi^{-\beta_m b} \Omega^{-\gamma_m c} \Omega^{-\gamma'_m c'} \cdots \log L_{a, b, c, c', \dots} \end{aligned}$$

Here the summation on the right-hand side of the equation is over the possible values of a, b, c, c', \dots . Once again, this translates to our Eq. (2).

Finally, consider the key use of the first orthogonality relation given by Corollary 3.5. In the special case where the common difference is a prime, Dirichlet expressed this by saying that we have

$$1 + \Omega^{h\gamma - \gamma_m} + \Omega^{2(h\gamma - \gamma_m)} + \cdots + \Omega^{(p-2)(h\gamma - \gamma_m)} = 0$$

except when $h\gamma - \gamma_m \equiv 0 \pmod{p-1}$, in which case the sum is equal to $p-1$. In the general case, he considered the sum $\frac{1}{h} \sum W_{\frac{1}{q^{h+h\rho}}}$,

...where the symbol \sum ranges over all primes q and W denotes the product of the sums taken over a, b, c, c', \dots or respectively over

$$\sum \Theta^{(h\alpha - \alpha_m)a}, \sum \Phi^{(h\beta - \beta_m)b}, \sum \Omega^{(h\gamma - \gamma_m)c}, \sum \Omega'^{(h\gamma' - \gamma'_m)c'}, \dots$$

Now one sees ...that the first of these sums is 2 or 0, corresponding to whether the congruence $h\alpha - \alpha_m \equiv 0 \pmod{2}$ or, equivalently, whether the congruence $q^h \equiv m \pmod{4}$ holds or not; that the second is $2^{\lambda-2}$ or 0 corresponding to whether the congruence $h\beta - \beta_m \equiv 0 \pmod{2^{\lambda-2}}$ or, equivalently, whether the congruence $q^h = \pm m \pmod{2^\lambda}$ holds or not; that the third is $(p-1)p^{\pi-1}$ or 0, corresponding to whether the congruence $h\gamma - \gamma_m \equiv 0 \pmod{(p-1)p^{\pi-1}}$ or, equivalently, whether the congruence $q^h \equiv m \pmod{p^\pi}$ holds or not, and so on; that therefore W always vanishes except when the congruence $q^h \equiv m \pmod{k}$ holds, in which case $W = K$.²³ (Dirichlet 1837b, p. 340)

Let us now summarize the differences between Dirichlet's presentation and contemporary ones. One salient difference is at the level of algebraic abstraction. The contemporary presentation developed the general notion of a character of an arbitrary finite abelian group, G , and showed that the characters themselves form a group, \widehat{G} , isomorphic to G itself. Dirichlet, in contrast, focused on a very particular group, $(\mathbb{Z}/k\mathbb{Z})^*$, the multiplicative group of units modulo k . His presentation showed an intimate familiarity with the structure of that group, and the explicit mapping from that group to the group of characters. These details are inessential in the modern presentation.

But there is another abstraction that distinguishes modern presentations from Dirichlet's, namely the willingness to treat characters as mathematical objects in their own right. This is, in part, facilitated by the algebraic abstraction: It is easier and more advantageous to treat characters as objects that transcend their representations when there are language and methods available that obviate the need to return to particular representations whenever there is real work to be done. But, the dependence goes both ways: The abstract algebraic methods cannot even be invoked until one is willing to consider characters as objects that can bear properties and algebraic structure.

It is the absence of these two forms of abstraction that puts Dirichlet's presentation in stark contrast with the modern one. Dirichlet did not define the notion of character at

²³ "...wo sich das Zeichen \sum auf die Primzahlen q erstreckt und W das Product der nach a, b, c, c', \dots resp. zwischen den angegebenen Grenzen zu nehmenden Summen:

$$\sum \Theta^{(h\alpha - \alpha_m)a}, \sum \Phi^{(h\beta - \beta_m)b}, \sum \Omega^{(h\gamma - \gamma_m)c}, \sum \Omega'^{(h\gamma' - \gamma'_m)c'}, \dots$$

bedeutet. Nun erseht man ...dass die erste dieser Summen 2 oder 0 ist, je nachdem die Congruenz $h\alpha - \alpha_m \equiv 0 \pmod{2}$, oder was dasselbe ist, die Congruenz $q^h \equiv m \pmod{4}$ stattfindet oder nicht stattfindet, das die zweite $2^{\lambda-2}$ oder 0 ist, je nachdem die Congruenz $h\beta - \beta_m \equiv 0 \pmod{2^{\lambda-2}}$, oder was dasselbe ist, die Congruenz $q^h = \pm m \pmod{2^\lambda}$ oder was dasselbe ist, die Congruenz $q^h \equiv m \pmod{4}$ stattfindet oder nicht stattfindet, das die dritte $(p-1)p^{\pi-1}$ oder 0 ist, je nachdem die Congruenz $h\gamma - \gamma_m \equiv 0 \pmod{(p-1)p^{\pi-1}}$ oder was dasselbe ist, die Congruenz $q^h \equiv m \pmod{p^\pi}$ stattfindet oder nicht stattfindet, u. s. w. das also W immer verschwindet, ausser wenn die Congruenz $q^h \equiv m \pmod{k}$ ist, in welchem Falle $W = K$ wird."

all, let alone general operations on characters; nor did he identify any of their general properties. Rather, characters only come into play as symbolic expressions in the construction of the L -functions, and their properties are derived in an ad hoc way as the proof proceeds. The characters are thus “intensional” entities: They are represented explicitly in terms of primitive roots of unity and generators of the corresponding residue group, and the only way of reasoning about them is in terms of this representing data.

In Dirichlet’s presentation, there is no quantifying over characters. Instead, one quantifies over their representations in terms of primitive roots of unity. Similarly, there is no direct notion of summing over characters: Dirichlet was happy to sum over finite sets of natural numbers and tuples of natural numbers, but where we would sum over a finite set of characters, Dirichlet instead summed over representations in terms of such tuples. It is notable that he presented the Euler product formula in (4) and (5) in terms of arbitrary primitive roots of unity and then explained that these “represent” or “contain” more particular equations that can then be summed over to obtain the desired result.

Finally, it is worth drawing attention to one further consequence of the different treatment of characters in the two presentations, namely the extent to which they make important dependences explicit. Because Dirichlet’s expressions depend on so much representational data, Dirichlet often suppressed details, to keep the expressions from getting unwieldy. Thus, he wrote ω^γ , suppressing the dependence of γ on n , where we would write $\chi(n)$ for the character χ corresponding to ω . This places a greater burden on the text of the proof, and the reader’s memory, to keep track of the relevant dependences, and, for example, the ranges of a summation. Moreover, the modern notation $L(s, \chi)$ makes it easy to track the dependence on the character χ , something that is lost in Dirichlet’s presentation. Thus, in the modern expression $\chi(m)L(s, \chi)$, the role of χ is clear in both terms in the product; in Dirichlet’s expression $\Omega^{-i\gamma_m} \log L_j$, the connection is buried in the definition of L_j .

6 Later presentations

In this section, we will examine proofs and discussions of Dirichlet’s theorem by Dedekind (Dirichlet 1863), de la Vallée Poussin (1897), Hadamard (1896), Kronecker (1901), and Landau (1909, 1927), as well as proofs of related results by Dirichlet (1840, 1841) and Weber (1882). The proofs of the original theorem all share the same structure, so we will focus on describing the similarities and differences between the presentations of characters and L -functions, rather than describing each in full.

6.1 Dirichlet

Over the next few years, Dirichlet extended the methods used to prove Theorem 3.1 in two different directions, first to consider primes represented by quadratic forms and second to consider arithmetic progressions in the complex numbers. We will consider each of these extensions briefly.

Regarding the first extension, some background on quadratic forms will be helpful. Number theory has long been concerned with questions as to which numbers can be represented by a given algebraic expression in which the variables are taken to range over the integers. It is easy to characterize the numbers that can be represented by a linear form $ax + b$ in one variable, and quadratic reciprocity addresses the problem of which numbers can be represented by a quadratic form $ax^2 + bx + c$. When it comes to quadratic forms in two variables, things become more difficult. Fermat's famous theorem that the odd prime numbers that can be represented by the form $x^2 + y^2$ are exactly the ones that are congruent to one modulo four is considered a gem of number theory. Euler proved this and other of Fermat's claims in the eighteenth century, and Lagrange later extended the theory of binary quadratic forms considerably. A central contribution of Gauss' *Disquisitiones Arithmeticae* is a complete classification of binary quadratic forms, whose study constitutes the bulk of that work.

Gauss showed that it suffices to characterize "primitive" quadratic forms $ax^2 + 2bxy + cy^2$, where the second coefficient is even, and a , b , and c have no factor in common. (Quadratic forms are also classified as "indefinite," "positive definite," and "negative definite," and in the discussion that follows it should be assumed that we are fixing our attention on one of these fixed kinds.) He called the value $D = b^2 - ac$ the *discriminant* of the form and showed how to assign to each primitive quadratic form of discriminant D a finite list of values of the form ± 1 which, roughly speaking, characterizes its behavior. He called these values the *characters* of the form and took the "genus" of a form to be the collection of all the primitive forms with the same discriminant and character. As we will see in Sect. 6.2, the use of name "character" for nonzero group homomorphisms to the complex numbers traces back to this terminology.

Dirichlet's Theorem 3.1 asserts that there are infinitely many primes represented by any linear form $kx + m$ where k and m have no common factor. In 1840, Dirichlet considered the analogous question as to whether infinitely many primes can be represented by a primitive quadratic form $ax^2 + 2bxy + cy^2$ and showed that this indeed holds in the special case where the discriminant $D = -p$ is a negative prime number, and p is congruent to 3 mod 4. As he noted, the proof relies on methods that are "essentially the same" ("im Wesentlichen ... übereinstimmend") (Dirichlet 1840, p. 98) as those used in his proof of the theorem on arithmetic progressions, although with a few new modifications. The result was proved in full generality by Weber (1882) in 1882, and de la Vallée Poussin (1897) also presented a proof in 1897.

In 1841, Dirichlet turned his attention to arithmetic progressions in the Gaussian integers and established that

...the expression $kt + l$, in which t denotes an indeterminate complex integer and k, l denote given numbers of this kind with no common factor, always contains infinitely many prime numbers."²⁴ (Dirichlet 1841, p. 509)

The proof begins with a discussion of modular arithmetic for the Gaussian integers and defines the notion of primitive roots and indices for complex moduli (Dirichlet 1841, pp. 512–524). Dirichlet then considered the product

²⁴ "...der Ausdruck $kt + l$, in welchem t eine unbestimmte complexe ganze Zahl und k, l gegebene solche Zahlen ohne gemeinschaftlichen Factor bezeichnen, immer unendlich viele Primzahlen enthält."

$$\Omega_n = \varphi^{\alpha_n} \varphi'^{\alpha'_n} \dots \psi^{\beta_n} \chi^{\gamma_n} \psi'^{\beta'_n} \chi'^{\gamma'_n} \dots \theta^{\delta_n} \eta^{\varepsilon_n}$$

where $\varphi, \varphi', \psi, \psi', \chi, \chi', \theta, \eta$ are all roots of unity and $\alpha_n, \alpha'_n, \beta_n, \beta'_n, \gamma_n, \gamma'_n, \dots, \delta_n, \varepsilon_n$ are indices with respect to n . This product is thus analogous to the characters that appear in his 1837 proof.

However, in contrast to the presentation in 1837, in 1841 Dirichlet explicitly noted that the products Ω_n enjoy a number of “important properties” (“wichtige Eigenschaften”) (Dirichlet 1841, p. 524). The properties he highlights are the following, where $\psi(k)$ is a quantity calculated from the prime decomposition of k , involving the norms of the prime factors and their exponents (Dirichlet 1841, 524–525):

1. $\Omega_{nn'} = \Omega_n \Omega_{n'}$.
2. $\Omega_{n'} = \Omega_n$ whenever $n' \equiv n \pmod{k}$.
3. $\sum \Omega_l = 0$ or $\sum \Omega_l = \frac{1}{4} \psi(k)$ depending on whether there is at least one root among the roots in Ω_l that is different to 1, or whether they are all equal.
4. $S\Omega_n = \frac{1}{4} \psi(k)$ or $S\Omega_n = 0$ depending on whether $n \equiv 1 \pmod{k}$ or $n \not\equiv 1 \pmod{k}$, where the sign “ S ” indicates a sum over all combinations of the roots that can occur in Ω .

Note that the last two are the key orthogonality relations, Theorem 3.4. Thus, in this presentation Dirichlet labeled the expressions that we now recognize as characters and isolated their properties. Moreover, he introduced a notation for summing over characters that is, interestingly, distinct from the notation for summation over natural numbers. The net effect is that the proof is considerably more modular than the proof of 1837.

6.2 Dedekind

In 1863, Dedekind published a book, *Vorlesungen über Zahlentheorie*, based on his notes taken from a course on number theory given by Dirichlet at Göttingen. After the lectures, he added nine “supplements,” or appendices, with material of his own. Supplement VI, in particular, contained a presentation of Dirichlet’s theorem. Dedekind extended the supplements in the second edition, published in 1871, to include his theory of algebraic ideals. That theory was, in turn, revised and expanded in the third and fourth editions, which appeared in 1879 and 1894, respectively. The earlier supplements, however, were not changed after the first edition.

Dedekind began his presentation with an overview of main steps of the proof. In particular, he proved the Euler product formula and obtained the series expansions for the L -functions and their logarithms. But the overview deals with a more general class of L -functions than we have considered so far:

The general proof of . . . [Dirichlet’s theorem] is based on the consideration of a class of infinite series of the form

$$L = \sum \psi(n),$$

where n runs through all positive integers and the real or complex function $\psi(n)$ satisfies the condition

$$\psi(n)\psi(n') = \psi(nn'),$$

...[and] we always assume that $\psi(1) = 1$.²⁵ (Dirichlet 1863, § 132)

He then went on to focus his attention on Dirichlet characters. He used the same expressions as Dirichlet to represent the characters, but whereas Dirichlet expressed each root of unity ω involved in the construction as the power Ω^i of a single primitive root of unity, Dedekind did not bother with this step. He did not explicitly refer to the relevant expressions as “characters,” but he introduced the notation $\chi(n)$ to denote their values, and pointed out that χ is completely multiplicative:

The numerator [of $\psi(n)$] $\chi(n) = \theta^\alpha \eta^\beta \omega^\gamma \omega'^{\gamma'} \dots$ has the characteristic property $\chi(n)\chi(n') = \chi(nn') \dots$ ²⁶ (Dirichlet 1863, §133, footnote)

Otherwise, however, Dedekind did not make use of the χ notation in his proof. And, like Dirichlet, Dedekind showed a reluctance to quantify over characters directly. He introduced the L -functions as

$$\prod \frac{1}{1 - \psi(q)} = \sum \psi(n) = L,$$

specifying that the product is to be taken over all primes not dividing the common difference, and the sum is to be taken over all natural numbers relatively prime to the common difference. Dedekind went on to note

...that the series can exhibit quite different behavior, depending on the roots of unity $\theta, \eta, \omega, \omega', \dots$ appearing in the expression for $\psi(n)$. Since these roots can have a, b, c, c', \dots values, respectively, the form L contains altogether

$$abcc' \dots = \varphi(k)$$

different particular series...²⁷ (Dirichlet 1863, §133)

²⁵ “Der allgemeine Beweis dieses Satzes ...stützt sich auf die Betrachtung einer Classe von unendlichen Reihen von der Form

$$L = \sum \psi(n),$$

wo der Buchstabe n alle ganzen positiven Zahlen durchlaufen muss, und die reelle oder complexe Function $\psi(n)$ der Bedingung

$$\psi(n)\psi(n') = \psi(nn')$$

genügt ...so nehmen wir immer an, dass $\psi(1) = 1$ ist.”

²⁶ “Der Zähler $\chi(n) = \theta^\alpha \eta^\beta \omega^\gamma \omega'^{\gamma'} \dots$ besitzt die charakteristischen Eigenschaften $\chi(n)\chi(n') = \chi(nn')$...”

²⁷ “Wir bemerken zunächst, dass diese Reihen je nach der Wahl der in dem Ausdrücke $\psi(n)$ vorkommenden Einheits-Wurzeln $\theta, \eta, \omega, \omega', \dots$ ein ganz verschiedenes Verhalten zeigen; da diese Wurzeln resp.

Note that the use of the word “contains” echoes Dirichlet’s language.

Dedekind then proceeded to divide the L -functions into classes. Given that the characters are defined in terms of a sequence of roots of unity, there are three distinct possibilities that can obtain for a given series L :

1. The roots of unity in the construction of the character occurring in L are all 1. There is only one such L -function, which is denoted L_1 .
2. The roots of unity in the construction of the character occurring in L are all real, i.e., are all ± 1 . L -functions that fall into this category are written as L_2 .
3. At least one of the roots of unity in the construction of the character occurring in L is imaginary. L -functions that fall into this category are written as L_3 .

Moreover, each L -function that falls into the third category has a conjugate: if $L_3 = \sum \frac{\theta^\alpha \eta^\beta \omega^\gamma \omega'^{\gamma'} \dots}{n^s}$, then there is a corresponding L -series in the same class, denoted by L'_3 , such that $L'_3 = \frac{\theta^{-\alpha} \eta^{-\beta} \omega^{-\gamma} \omega'^{-\gamma'} \dots}{n^s}$.

Let us now see how Dedekind’s notation plays out when it comes to summing over the characters. When proving that the L -functions corresponding to a complex character have a finite nonzero value as s tends to infinity, he obtained the following equation:

$$\begin{aligned} \varphi(k) \left(\sum \frac{1}{q^s} + \frac{1}{2} \sum \frac{1}{q^{2s}} + \dots + \frac{1}{\mu} \sum \frac{1}{q^{\mu s}} + \dots \right) \\ = \log L_1 + \sum \log(L_2) + \sum \log(L_3 L'_3), \end{aligned} \quad (6)$$

where, on the left hand side, the successive sums are over all the primes q not dividing k which satisfy the successive conditions $q \equiv 1 \pmod{k}$, $q^2 \equiv 1 \pmod{k}$, etc. On the right hand side the first sum is over all series L_2 of the second class, and the second sum is over all conjugate pairs $L_3 L'_3$ of series of the third class.²⁸ (Dirichlet 1863, §136)

In a sense, Dedekind took the summations to range not over the characters or their representations, but over the L -functions themselves. Later, when dealing with these

Footnote 27 Continued

a, b, c, c', \dots verschiedene Werthe haben können, so sind in der Form L im Ganzen

$$abcc' \dots = \varphi(k)$$

verschiedene besondere Reihen enthalten ...”

28 ..

$$\varphi(k) \left(\sum \frac{1}{q^s} + \frac{1}{2} \sum \frac{1}{q^{2s}} + \dots + \frac{1}{\mu} \sum \frac{1}{q^{\mu s}} + \dots \right) = \log L_1 + \sum \log(L_2) + \sum \log(L_3 L'_3),$$

wo auf der linken Seite das erste, zweite Summenzeichen u.s.f. sich auf alle die in k nicht aufgehenden Primzahlen q bezieht, welche resp. den Bedingungen $q \equiv 1$, $q^2 \equiv 1 \pmod{k}$ u.s.f. Genüge leisten; auf der rechten Seite bezieht sich das erste Summenzeichen auf alle Reihen L_2 der zweiten Classe, das zweite auf alle verschiedenen Paare $L_3 L'_3$ conjugirter Reihen dritter Classe.” We have added the equation numbers in this quotation and the next, for later reference.

sums, he came closer to Dirichlet's presentation, in that the indices of the summation range over the sequences of roots that determine the characters. However, at times he introduced convenient abbreviations. For example, given a particular collection of roots of unity $\theta, \eta, \omega, \omega', \dots$, Dedekind denoted $\theta^{-\alpha_1} \eta^{-\beta_1} \omega^{-\gamma_1} \omega'^{-\gamma'_1} \dots$ by χ , where $\alpha_1, \beta_1, \gamma_1, \gamma'_1$ stand for the indices of the first term of the progression, m . Note that while Dedekind used the notation $\chi(n)$ in a footnote, he explicitly called χ , as defined here, a *value*. (In fact, it plays the role of $\overline{\chi(m)}$ in the presentation of Sect. 3.3.) Moreover, when he used the symbol in a summation, Dedekind was explicit that the summation ranges not over χ , but rather the roots of unity involved in the definition:

The summation of all products $\chi \log L$ therefore gives the result

$$\varphi(k) \left(\sum \frac{1}{q^s} + \frac{1}{2} \sum \frac{1}{q^{2s}} + \frac{1}{3} \sum \frac{1}{q^{3s}} + \dots \right) = \sum \chi \log L, \quad (7)$$

where the successive sums on the left hand side are over all primes q satisfying the successive conditions $q \equiv m, q^2 \equiv m, q^3 \equiv m \pmod{k}$ etc., while the sum on the right hand side is over all $\varphi(k)$ different root systems $\theta, \eta, \omega, \omega', \dots$ ²⁹ (Dirichlet 1863, §137)

In many ways, Dedekind did not stray far from Dirichlet's presentation. For the most part, his treatment of the characters was intensional, in the sense that the arguments rely on the particular representations of the characters. In other words, operations on the characters are described in terms of their representations, rather than their values, and, like Dirichlet, he viewed summations as ranging over these representations in Eqs. (6) and (7). Similarly, he classified the characters depending on the roots involved in their representation, rather than their values. Nor did he take the expression L , for the L -functions, to depend on the characters themselves. Rather, he characterized them as expressions that behave differently "depending on the roots of unity" appearing in their construction. The L notation in particular does nothing to signal this dependence.

Nonetheless, he did take some key steps toward viewing the characters abstractly. To start with, Dedekind went out of his way to isolate the characters as independent of the L -series in which they appear and, in particular, flagged them as entities satisfying certain key properties. And, at least at times, he characterized summations as ranging over the L -functions themselves, hinting at a new level of abstraction.

The use of the term "character" in the modern sense can be traced to the long Supplement XI in the 1879 edition of the Dirichlet–Dedekind *Vorlesungen*, which

²⁹ "Die Summation aller Producte $\chi \log L$ giebt daher das Resultat

$$\varphi(k) \left(\sum \frac{1}{q^s} + \frac{1}{2} \sum \frac{1}{q^{2s}} + \frac{1}{3} \sum \frac{1}{q^{3s}} + \dots \right) = \sum \chi \log L,$$

wo auf der linken Seite das erste, zweite, dritte Summenzeichen u.s.f. sich auf alle Primzahlen q bezieht, welche resp. den Bedingungen $q \equiv m, q^2 \equiv m, q^3 \equiv m \pmod{k}$ u.s.f. genügen, während das Summenzeichen auf der rechten Seite sich auf die sämmtlichen $\varphi(k)$ verschiedenen Wurzel-Systeme $\theta, \eta, \omega, \omega', \dots$ "

contains a presentation of Dedekind’s theory of ideals in algebraic number fields. As noted in Sect. 6.1, Gauss assigned “characters” to quadratic forms in order to classify their behavior. Already in the first edition of the *Vorlesungen*, Dedekind used Gauss’ terminology in Supplement IV, “Genera of quadratic forms.” In 1879, however, Dedekind went further by showing that Gauss’ classification could be understood in terms of his theory of ideals. Specifically, he showed that there is a correspondence between genera of quadratic forms and equivalence classes of ideals in an associated quadratic extension of the rationals. Moreover, there is a group structure on the ideals, and Gauss’ characters correspond to characters on that group, in the modern sense. This explains the quotation in Sect. 2: Dedekind had simply adopted Gauss’ terminology to characterize the homomorphisms from groups of ideals to the (nonzero) complex numbers more generally.

As we noted in Sect. 2.1, in 1882, Weber (1882), building on Dedekind’s work, extended the notion of a “character” to arbitrary finite abelian groups. But, although Dedekind revised his theory of ideals substantially in the third and fourth editions of the *Vorlesungen*, he did not revise any of the supplements that appeared in the first edition. In particular, he did not take advantage of the opportunity to go back to introduce the modern notion of a character in his presentation of Dirichlet’s proof, although he clearly could have done so in the later editions.

6.3 Weber

The paper (Weber 1882) is devoted to a proof of the fact that any primitive quadratic form represents infinitely many primes, the result that Dirichlet had proved in 1840 in a special case. That paper, however, was the first to present the modern notation of a character on a finite abelian group in full generality. Weber defined such a group as consisting of h elements “of any type” (“irgend welcher Art”) satisfying the usual group laws. He then stated the now-familiar structure theorem:

In a finite abelian group of order h one can always choose elements $\Theta_1, \Theta_2, \dots, \Theta_\nu$ of order n_1, n_2, \dots, n_ν so that each element Θ of G can be expressed uniquely by the form

$$\Theta_1^{s_1} \Theta_2^{s_2} \dots \Theta_\nu^{s_\nu},$$

where s_1, s_2, \dots, s_ν are chosen from complete residue systems with respect to the moduli n_1, n_2, \dots, n_ν .³⁰ (Weber 1882, pp. 306–307)

³⁰ “In einer Abel’schen Grupper G von Grade h kann man stets die Elemente $\Theta_1, \Theta_2, \dots, \Theta_\nu$ von den Graden n_1, n_2, \dots, n_ν so auswählen, dass in der Form

$$\Theta_1^{s_1} \Theta_2^{s_2} \dots \Theta_\nu^{s_\nu}$$

jedes Element Θ von G und jedes nur einmal enthalten ist, wenn s_1, s_2, \dots, s_ν je einem vollständigen Restsystem nach den Moduln n_1, n_2, \dots, n_ν entnommen werden.”

Weber called the sequence of values $\Theta_1, \Theta_2, \dots, \Theta_\nu$ a *basis* of the group and used them to define the characters as follows:

If we assign to the ν elements $\Theta_1, \Theta_2, \dots, \Theta_\nu$ of such a basis ν roots of unity $\omega_1, \omega_2, \dots, \omega_\nu$, of order n_1, n_2, \dots, n_ν respectively, then each element $\Theta = \Theta_1^{s_1} \Theta_2^{s_2} \dots \Theta_\nu^{s_\nu}$ of the group also corresponds to a particular h^{th} root of unity

$$\omega = \omega_1^{s_1} \omega_2^{s_2} \dots \omega_\nu^{s_\nu}.$$

We denote this root of unity by $\chi(\Theta)$ and call it the *character* of the element.³¹ (Weber 1882, p. 307)

Weber went on to point out that this gives rise to h distinct characters and that each such character χ satisfies the equation

$$\chi(\Theta)\chi(\Theta') = \chi(\Theta\Theta').$$

Moreover, this last condition provides an exact characterization:

If, conversely, $\chi(\Theta)$ is a uniquely determined function of Θ which satisfies [the equation above], then it is necessarily contained among these h characters.³² [*ibid.*]

Thus, Weber's paper provides us not only with the modern notion of a character defined on an arbitrary finite abelian group, but also with the modern understanding that such characters constitute an instance of the function concept.

There are some small differences between Weber's presentation and ours. In his presentation of the second orthogonality relation, Weber used ellipses rather than summation notation to sum over the characters:

For each element Θ :

$$\chi_1(\Theta) + \chi_2(\Theta) + \dots + \chi_h(\Theta) = 0,$$

except for the identity element Θ_0 , for which we have

$$\chi_1(\Theta_0) + \chi_2(\Theta_0) + \dots + \chi_h(\Theta_0) = h.$$

³¹ "Ordnet man den ν Elementen $\Theta_1, \Theta_2, \dots, \Theta_\nu$ einer solchen Basis ν Einheitswurzeln $\omega_1, \omega_2, \dots, \omega_\nu$ von den Graden n_1, n_2, \dots, n_ν , so entspricht auch jedem Element $\Theta = \Theta_1^{s_1} \Theta_2^{s_2} \dots \Theta_\nu^{s_\nu}$ der Gruppe eine bestimmte h^{te} Einheitswurzel ω nach der Vorschrift

$$\omega = \omega_1^{s_1} \omega_2^{s_2} \dots \omega_\nu^{s_\nu}.$$

Wir bezeichnen diese Einheitswurzel mit $\chi(\Theta)$, und nennen dieselbe den *Charakter* des Elementes Θ ."

³² Ist umgekehrt $\chi(\Theta)$ eine durch das Element Θ eindeutig bestimmte Function, welche der Bedingung ... genügt, so ist dieselbe nothwendig unter diesen h Charakteren enthalten.

(Weber 1882, p. 308)³³

This implicitly assumes that the h characters are enumerated χ_1, \dots, χ_h , though the enumeration is arbitrary. Moreover, Weber provided the explicit construction of the set of characters before the extensional characterization, whereas modern presentations usually provide the extensional characterization first. But these points are minor, and, otherwise, his presentation is little different from the one in Sect. 3.2.³⁴ Even his fairly modern presentation harks back to an intensional view, however. After introducing analogues of the Euler product formula and the result of taking logarithms of both sides, he emphasized, echoing the language used by Dirichlet and Dedekind:

Each of the formulas ...represents h different formulas, corresponding to the h different characters $\chi_1, \chi_2, \dots, \chi_h$.³⁵ (Weber 1882, p. 315)

He then characterized the analogue of Eq. (2) in Sect. 3.4 as the result of “the addition of all the formulas” (“die Addition der sämtlichen Formeln”), rather than the addition of the corresponding values.

6.4 De la Vallée-Poussin

More than two decades later, Charles Jean de la Vallée-Poussin gave another presentation of Dirichlet’s theorem in a 1895/6 paper entitled “Démonstration simplifiée du Théorème de Dirichlet sur la progression arithmétique,” and again in sections from his 1897 book *Recherches analytiques sur la théorie des nombres premiers*. He introduced the characters in much the same way that Dirichlet and Dedekind did, namely, via an explicit construction in terms of primitive roots and roots of unity. Like Dedekind and Weber, he used the symbol χ for the characters. Like Weber, he distinguished between the $\varphi(M)$ characters modulo M using subscripts, writing $\chi_1, \chi_2, \dots, \chi_{\varphi(M)}$.

De la Vallée-Poussin went on to subject the characters to a thorough study. Whereas Dirichlet and Dedekind divided the L -functions (or, perhaps more precisely, the corresponding expressions) into three categories based on their representations, de la Vallée-Poussin based the categorization on the characters themselves.

³³ “Für jedes Element Θ ist:

$$\chi_1(\Theta) + \chi_2(\Theta) + \dots + \chi_h(\Theta) = 0.$$

ausgenommen für das Hauptelement Θ_0 , für welches

$$\chi_1(\Theta_0) + \chi_2(\Theta_0) + \dots + \chi_h(\Theta_0) = h.$$

³⁴ Mackey’s historical survey (Mackey 1980) of the history of harmonic analysis includes a very helpful and informative overview of the history of character-theoretic ideas in number theory.

³⁵ “Jede der Formeln ...repräsentirt h verschiedene Formeln, entsprechend den h verschiedenen Charakteren $\chi_1, \chi_2, \dots, \chi_h$.”

The following description applies to the simplest case, where the modulus M is a prime:

One calls the character that corresponds to the root $+1$ the *principal character*; it is equal to unity for all numbers n . Apart from the principal character, there is only one which is real for all numbers n : it corresponds to the root (-1) and is equal to ± 1 depending on the number n .

We call all the other characters *imaginary characters*, though they may have a real value for some particular numbers. Their modulus is always equal to unity.³⁶ (de la Vallée Poussin 1897, p. 19)

Thus, de la Vallée-Poussin divided the characters into three classes:

1. the class consisting solely of the principal character;
2. the class consisting of all real characters (in this case, there is only one character that is real for all numbers, corresponding to the root -1); and
3. The class consisting of all other characters, called the imaginary characters.

Notice that his categorization includes both intensional and extensional characterizations; that is, he characterized the characters in each class in terms of both the values they take and the roots involved in their construction.

In his study of the characters in his 1896 paper, de la Vallée Poussin listed a number of “very important relations” (“relations très importantes”). The first is that $\chi(n)\chi(n') = \chi(nn')$ for every character χ , and n and n' . The second is that for a given character χ modulo M , and any n and n' , if $n \equiv n' \pmod{M}$, then $\chi(n) = \chi(n')$. The third and fourth are the first and second orthogonality relations, respectively. In order to present the second orthogonality relation, de la Vallée-Poussin introduced a new symbol, S , to denote summation over the characters, just as Dirichlet did in 1841:³⁷

³⁶ “On appelle *caractère principal* celui qui correspond à la racine $+1$; il est égal à l’unité pour tous les nombres n . En dehors du caractère principal, il n’y en a qu’un seul qui soit réel pour tous les nombres n : il correspond à la racine (-1) et est égal à ± 1 suivant le nombre n .

Nous donnerons à tous les autres caractères le nom de *caractère imaginaires*, quoiqu’ils puissent avoir une valeur réelle pour certains nombres particuliers. Leur module est toujours égal à l’unité.”

³⁷ “Considérons ... la somme étendue à tous les caractères, c’est-à-dire à tous les systèmes de racines

$$S_{\chi} \chi(n) = S_{\omega} \omega_1^{v_1} \omega_2^{v_2} \dots$$

... Pour tout nombre n , la somme étendue à la totalité des caractères

$$S_{\chi} \chi(n) = 0,$$

à la seule exception près du cas où

$$n \equiv 1 \pmod{M},$$

car alors tous les indicateurs sont nuls et l’on a

$$S_{\chi} \chi(n) = \varphi(M)."$$

Consider ...the sum extending over all the characters, that is to say over all the systems of roots

$$S_{\chi} \chi(n) = S_{\omega} \omega_1^{\nu_1} \omega_2^{\nu_2} \dots$$

...For every number n , the sum extending over the totality of characters satisfies

$$S_{\chi} \chi(n) = 0,$$

the only exception being the case where

$$n \equiv 1 \pmod{M},$$

because then all the indices are zero and one has

$$S_{\chi} \chi(n) = \varphi(M).$$

[de la Vallée Poussin, pp. 14–15]

Although his notation is different from that of Dirichlet's original notation and Dedekind's, his proof of Theorem 3.1, like theirs, relies on an explicit calculation based on the construction of the characters, in contrast to the modern proof sketched in Sect. 3.2.

When introducing the L -functions in his 1896 paper, de la Vallée-Poussin simply described them in terms of their equivalent expressions as a sum and product. In 1897, however, he adopted a functional notation:

We define the function $Z(s, \chi \bmod M)$, for $\mathcal{R}(s) > 1$, by the absolutely convergent expressions:

$$Z(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod \left(1 - \frac{\chi(q)}{q^s} \right)^{-1}$$

where n designates successively all the integers prime to M and q all the prime numbers not dividing M .³⁸ (de la Vallée Poussin 1897, p. 56)

Aside from the choice of the letter Z instead of the letter L , we have finally arrived at the modern notation. But, like Dirichlet and Dedekind, he was reticent to quantify

³⁸ "Nous définirons la fonction $Z(s, \chi \bmod M)$, pour $\mathcal{R}(s) > 1$, par les expressions absolument convergentes

$$Z(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod \left(1 - \frac{\chi(q)}{q^s} \right)^{-1}$$

où n désigne successivement tous les nombres entiers premiers à M et q tous les nombres premiers qui ne divisent pas M ."

over the characters, using language that is eerily reminiscent of theirs. For example, in his 1897 work, he wrote:

...one finds the *fundamental equation*

$$(E) \dots - \lim_{s=1} (s-1) \frac{Z'(s, \chi)}{Z(s, \chi)} = \lim_{s=1} (s-1) \sum_q \chi(q) \frac{lq}{q^s},$$

and this equation (E) represents in reality $\varphi(M)$ distinct ones, which result from exchanging the characters amongst themselves.³⁹ (de la Vallée Poussin 1897, p. 65).

Nonetheless, there are a number of important respects in which de la Vallée-Poussin's presentation is close to the modern one. To start with, characters are treated as objects of study in their own right, bearing their own properties and relations. Moreover, they are classified extensionally, although this classification is, at the same time, related to properties of their representations. They are also represented notationally as arguments to the L -functions. Finally, de la Vallée-Poussin used a summation notation with an index ranging over the characters, not their representing data. The fact that he went out of his way to use one symbol, S , for summation of characters and the usual Σ symbol for summation over sets of natural numbers hints that he does not conceive of these as precisely the same mathematical operation, although they have similar properties.

6.5 Hadamard

For each real number x , let $\pi(x)$ denote the number of primes less than x . Around the turn of the century, both Legendre and Gauss conjectured that $\pi(x)$ is asymptotic to $x/\ln(x)$, in the sense that their ratio, $\pi(x) \ln x/x$, approaches 1 as x approaches infinity. This fact, now known as the “prime number theorem,” was finally proved by both de la Vallée-Poussin and Jacques Hadamard, working independently, in 1896. De la Vallée-Poussin's proof was first published in the *Annales de la Société Scientifique de Bruxelles*, but also appeared in the 1897 book we discussed in the last section. Both Hadamard and de la Vallée-Poussin also obtained a generalization of the prime number theorem to arithmetic progressions, which says that if m is relatively prime to k , the number of primes less than x that are congruent to m modulo k is asymptotic to $(1/\varphi(k)) \cdot (x/\ln x)$. This is, of course, a generalization of Dirichlet's theorem, since it implies that there are infinitely many primes congruent to m modulo k .

Hadamard's 1896 paper was titled “Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques” and provided a number of results concerning

³⁹ “...on trouve l'équation fondamentale

$$(E) \dots - \lim_{s=1} (s-1) \frac{Z'(s, \chi)}{Z(s, \chi)} = \lim_{s=1} (s-1) \sum_q \chi(q) \frac{lq}{q^s},$$

et cette équation (E) en représente en réalité $\varphi(M)$ distinctes par l'échange des caractères entre eux.”

L -functions and characters. He introduced characters in the same way as Dirichlet, Dedekind, and de la Vallée-Poussin, namely, by a construction in terms of roots of unity and primitive elements. Like Weber and de la Vallée-Poussin, he distinguished between the characters modulo k notationally by making use of subscripts, writing $\psi_v(n)$ where v runs from 1 to $\varphi(k)$. He defined the L -functions as follows:

$$L_v(s) = \sum_{n=1}^{\infty} \frac{\psi_v(n)}{n^s}.$$

The notation $L_v(x)$, in contrast to de la Vallée-Poussin’s $Z(s, \chi)$, does not indicate an explicit dependence on the character, though the subscript provides a link. As did his predecessors, Hadamard classified the L -functions into three different classes, but his classification was intensional, like Dirichlet’s and Dedekind’s, referring to the roots of unity used in the construction of the characters.

When it came to summing over the characters, Hadamard, unlike Dirichlet, Dedekind, or de la Vallée-Poussin, let the index of the summation range over the *subscripts* described above.

The fundamental equation that Dirichlet used in the demonstration of his theorem is

$$\sum_v \frac{\log L_v(s)}{\psi_v(m)} = \varphi(k) \left(\sum \frac{1}{q^s} + \frac{1}{2} \sum' \frac{1}{q^{2s}} + \frac{1}{3} \sum'' \frac{1}{q^{3s}} + \dots \right),$$

where m is some integer prime to k , and where, among the signs $\sum, \sum', \sum'', \dots$, the first ranges over the prime numbers q such that $q \equiv m \pmod{k}$, the second ranges over the primes numbers q such that $q^2 \equiv m \pmod{k}$, etc.⁴⁰ (Hadamard 1896, p. 209)

Once again, this is comparable to the modern formulation in Eq. (2).

Hadamard’s presentation falls somewhere between those of Dirichlet and Dedekind, on the one hand, and that of de la Vallée-Poussin, on the other. His treatment of characters was, for the most part, intensional. But, in contrast to Dirichlet and Dedekind, he left representational data out of the key summations, thereby eliminating unnecessary clutter. However, unlike de la Vallée-Poussin, he did not go so far as to characterize these as summations over the characters themselves. Rather, he introduced natural number indices, v , labeling the characters, and then took the variable of summation to range over those.

⁴⁰ “L’équation fondamentale utilisée par Dirichlet pour la démonstration de son théorème, est

$$\sum_v \frac{\log L_v(s)}{\psi_v(m)} = \varphi(k) \left(\sum \frac{1}{q^s} + \frac{1}{2} \sum' \frac{1}{q^{2s}} + \frac{1}{3} \sum'' \frac{1}{q^{3s}} + \dots \right)$$

où m est un entier quelconque premier avec k et où les signes $\sum, \sum', \sum'', \dots$ s’étendent, le premier aux nombres premiers q tels que $q \equiv m \pmod{k}$, le second aux nombres premiers q tels que $q^2 \equiv m \pmod{k}$, etc.”

6.6 Kronecker

Between 1863 and 1891, Leopold Kronecker lectured at the University of Berlin on a range of subjects, including number theory, algebra, and the theory of determinants. After Kronecker's death, his student, Kurt Hensel, edited the five volumes of his collected works, which were published between 1895 and 1930. Hensel also took it upon himself to work Kronecker's copious lecture notes and course material into two textbooks, *Vorlesungen über Zahlentheorie* and *Vorlesungen über die Theorie der Determinanten*, which he published in 1901 and 1903, respectively. The first of these closes with a proof of Dirichlet's theorem, which we will discuss here.

Kronecker, in fact, wrote his doctoral dissertation on algebraic number theory under Dirichlet's supervision, completing it in 1845. Kronecker insisted that mathematics should maintain a clear focus on symbolic representations and algorithms, a commitment that is evident throughout his work. Avoiding talk of "arbitrary" functions, real numbers, and so on, Kronecker focused instead on the construction of algebraic systems and explicit algorithms for calculating with these algebraic representations. For example, his article "Grundzüge einer arithmetischen Theorie der algebraischen Grössen" (Kronecker 1882) provides means of carrying out operations on systems of algebraic integers in finite extensions of the rationals. Similarly, "Ein Fundamentalsatz der allgemeinen Arithmetik" (Kronecker 1887) provides an explicit construction of a splitting field for any polynomial with integer coefficients, and he viewed this as filling a gap in Galois' work.⁴¹

Kronecker's approach to number theory had a similar orientation. As Hensel put it in the introduction to *Vorlesungen über Zahlentheorie*:

He believed that one can and must in this domain formulate each definition in such a way that its applicability to a given quantity can be assessed by means of a finite number of tests. Likewise, an existence proof for a quantity is to be regarded as entirely rigorous only if it contains a method by which that quantity can actually be found. Kronecker was far from being one to completely reject a definition or proof that does not meet these highest requirements, but he believed that there was then something missing, and he held that completing it along these lines is an important task, by which our knowledge is furthered in an important sense.⁴² (Kronecker 1901, p. vi)

⁴¹ See Edwards (1980), Edwards (1989), Edwards (2007), Edwards (2009) for a discussion of these works, and Kronecker's mathematics more generally.

⁴² "Er meinte, man könne und man müsse in diesem Gebiete eine jede Definition so fassen, daß durch eine endliche Anzahl von Versuchen geprüft werden kann, ob sie auf eine vorgelegte GröÙe anwendbar ist oder nicht. Ebenso wäre ein Existenzbeweis für eine GröÙe erst dann als völlig streng anzusehen, wenn er zugleich ein Method enthalte, durch welche die GröÙe, deren Existenz bewiesen werde, auch wirklich gefunden werden kann. Kronecker war weit davon entfernt, eine Definition oder einen Beweis vollständig zu verwerfen, der jenen höchsten Anforderungen nicht entsprach, aber er glaubte, daß dann eben noch etwas fehle, und er hielt eine Ergänzung nach dieser Richtung hin für eine wichtige Aufgabe, durch die unsere Erkenntnis in einem wesentlichen Punkte erweitert würde." We have modified and extended a translation due to Stein (1988, p. 250).

Kronecker’s proof of Dirichlet’s theorem is a focal point of the book, in that it brings together methods and ideas developed throughout the entire work. Here is Hensel’s characterization:

[The book] closes with the proof of the famous theorem that any arithmetic sequence, whose first term and common difference are relatively prime, contains infinitely many prime numbers. But Kronecker completed Dirichlet’s proof of this theorem in a significant sense, in that he proved that one can determine, for an arbitrarily large number μ , a larger number $\bar{\mu}$, so that in the interval $(\mu \cdots \bar{\mu})$ one is sure to find a prime number of the required form. This nice supplement to that famous proof is a fruit of the higher demands, mentioned above, that Kronecker placed on arithmetic proofs. And here it seems, in fact, that with this improvement of Dirichlet’s theorem, nothing by way of simplicity or transparency has been lost.⁴³ (Kronecker 1901, p. viii)

The interaction between analytic and number-theoretic methods is a central theme of the *Vorlesungen*, which opens with a discussion of Gauss’ distinction between the fields of number theory, algebra, and analysis, and argues that they cannot be cleanly separated. For example, Kronecker praised Leibniz’ characterization of π in terms of the series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}$$

as providing a definition of π of a “fully number-theoretic character” (“durchaus zahlentheoretischem Charakter”).

What these examples teach us holds, more generally, for all the definitions of analysis. These always lead back to the integers and their properties, and from that entire branch of mathematics, only the concept of a limit has so far remained foreign. Arithmetic cannot be separated from analysis, which has freed itself from its original source, geometry, and has developed independently, on free soil. Even less so, as Dirichlet has succeeded in obtaining the deepest and most beautiful results in arithmetic by combining the methods of the two disciplines.⁴⁴ (Kronecker 1901, pp. 4–5)

⁴³ “...schließt mit dem Beweise des berühmten Satzes, daß jede arithmetische Reihe, deren Anfangsglied und Differenz teilerfremd sind, unendlich viele Primzahlen enthält; aber Kronecker vervollständigt den Dirichletschen Beweis dieses Satzes in einem wesentlichen Punkte, indem er nachweist, daß man für jede beliebige groß anzunehmende Zahl μ eine größere Zahl $\bar{\mu}$ so bestimmen kann, daß in dem Intervalle $(\mu \cdots \bar{\mu})$ sich sicher eine Primzahl der verlangten Form befindet. Dies schöne Ergänzung jenes berühmten Beweises ist eine Frucht der oben erwähnten höheren Forderungen, welche Kronecker an arithmetische Beweise stellte, und hier scheint es in der That, daß durch diese Verbesserung der Dirichletsche Beweis nichts an Einfachheit und Durchsichtigkeit verloren hat.”

⁴⁴ “Was uns diese Beispiele lehren, ist nun maßgebend für alle Definitionen der Analysis überhaupt. Dieselben führen stets auf die ganzen Zahlen und ihre Eigenschaften zurück, und es ist von dem ganzen Gebiete des letztgenannten Zweiges der Mathematik der einzige Begriff des limes oder der Grenze der Zahlentheorie bisher fremd geblieben. Gegen die Analysis also, die sich von ihrer ursprünglichen Quelle, der Geometrie, befreit und auf freiem Boden selbständig entwickelt hat, kann die Arithmetik nicht abgegrenzt

This interplay comes to the fore in the proof of Dirichlet's theorem. The text reports that Kronecker's version of the proof, in the case where the common difference is prime, was worked out in the lectures he gave in the winter semester of 1875/1876, whereas the general case was presented in the winter semester of 1886/1887 (Kronecker 1901, p. 442). Fixing a modulus m and an r relatively prime to m , recall that Kronecker aimed to provide, for a given number μ , an explicit upper bound $\bar{\mu}$ on the integers one has to consider to find a prime number greater than μ and congruent to r modulo m . When it came to the analytic part of the proof, Kronecker explained that obtaining the desired bound is reduced to obtaining a positive lower bound on a certain analytic series that arises in the proof. He wrote:

For the ambiguous [real] characters, Dirichlet's proof meets this requirement. But his methods are not sufficient to do the same for the series corresponding to the complex characters.⁴⁵ (Kronecker 1901, p. 481)

He went on:

Generally speaking, this is a special case of the problem of finding, given a well-defined nonzero number, a bound, above which it necessarily lies. This is not as easy as it seems at first glance; indeed, in some circumstances, the problem can count among the thorniest questions known to science.⁴⁶ (Kronecker 1901, pp. 481–482)

Kronecker took the opportunity to clarify the methodological stance toward analysis that is appropriate to these issues. Even though, in his work, he avoided the notion of an "arbitrary" real number, he recognized the importance of understanding particular number systems in analytic terms. For example, given a symbolic representation of a real number, one may wish to compute rational approximations. But knowing that a nonnegative real number is nonzero is not the same as having a positive, rational lower bound on its values.⁴⁷ Kronecker mentioned the problem of bounding a nonzero determinant away from zero as an example of a problem that is generally difficult. He also noted that, given two convergent series, it can be difficult to determine which one

Footnote 44 Continued

werden, um so weniger, als es *Dirichlet* gelungen ist, grade die schönsten und tiefliegenden arithmetischen Resultate durch die Verbindung der Methoden beider Disciplinen zu erzielen."

⁴⁵ "Für die ambigen Charaktere erfüllt der Beweis von Dirichlet auch diese Forderung, dagegen reichen seine Methoden nicht aus, um dasselbe auch für die Reihen zu leisten, welche den complexen Charakteren entsprechen."

⁴⁶ "Überhaupt ist das hier in einen speziellen Falle sich darbietende Problem, für eine von Null verschiedene wohldefinierte Zahlgröße eine Grenze zu finden, *über* der sie notwendig liegen muß, nicht so einfach, als es auf den ersten Blick erscheint, vielmehr kann diese Aufgabe unter Umständen eine der heikelsten Fragen sein, die die Wissenschaft kennt."

⁴⁷ In modern terms, one can obtain such a lower bound by computing rational approximations until one obtains one that is sufficiently accurate to bound the number away from zero. Thus, the statement "if $r \neq 0$, then $|r| > 0$ " was accepted by the Russian school of constructive mathematics in the 1950s and 1960s. This implication is equivalent to "Markov's principle," which is, however, rejected by strict constructivists. See, for example, Troelstra and Dalen (1988).

has a greater value. These facts are now quite familiar in constructive and computable analysis.⁴⁸

Kronecker’s treatment of characters provides an interesting combination of Dirichlet’s approach and modern ones. Fixing a modulus m , Kronecker, like Dirichlet, provided a fully explicit description of the characters modulo m in terms of primitive elements of the powers of primes giving m and primitive roots of unity. However, with judicious choice of notation, he was at the same time able to suppress extraneous detail. For example, fixing primitive elements $\gamma, \gamma_0, \gamma_1, \dots, \gamma_g$, one can express any element r relatively prime to m in the form

$$r \equiv \gamma^\rho \gamma_0^{\rho_0} \cdots \gamma_g^{\rho_g} \bmod m.$$

But then one can “package” the representing values $(\rho, \rho_0, \rho_1, \dots, \rho_g)$ as a tuple, the “index system of r ,” denoted $\text{Indd } r$. For each of the cyclic groups in the decomposition of the group of units modulo m , suppose we also choose corresponding roots of unity, $\omega, \omega_0, \omega_1, \dots, \omega_g$.⁴⁹

Now let r be a unit modulo m , and $\text{Indd } r = (\rho, \rho_0, \rho_1, \dots, \rho_g)$; we now assign to r the root of unity:

$$\Omega(r) = (-1)^\rho \omega_0^{\rho_0} \omega_1^{\rho_1} \cdots \omega_g^{\rho_g}$$

which we call a *character* of r , since the index system $(\rho, \rho_0, \rho_1, \dots, \rho_g)$, and hence $\Omega(r)$, are uniquely determined.⁵⁰ (Kronecker 1901, p. 444)

We obtain *all* the possible characters by fixing primitive roots of unity $\omega, \omega_0, \dots, \omega_g$, so that every tuple of roots can be represented in the form

$$\omega^k, \omega_0^{k_0}, \omega_1^{k_1}, \dots, \omega_g^{k_g},$$

where the k and k_i ’s are less than the cardinality of the corresponding cyclic group.

When there is no fear of misunderstanding, we will, in the following, denote the underlying exponent system (k, k_0, k_1, \dots) by (k) for short, and we will denote the corresponding character simply by

$$\Omega^{(k)}(r).$$

⁴⁸ See, for example, Troelstra and Dalen (1988).

⁴⁹ As above, the powers of two require special treatment. In the discussion, Kronecker assumes that m is divisible by 8, in which case the units modulo that power of two form a product of two cyclic groups; the values of ρ and ρ_0 are the indices in those two groups, and ω and ω_0 are the corresponding roots of unity.

⁵⁰ Es sie nun r eine Einheit modulo m , und $\text{Indd } r = (\rho, \rho_0, \rho_1, \dots, \rho_g)$; ordnen wir r jetzt die Einheitswurzel:

$$\Omega(r) = (-1)^\rho \omega_0^{\rho_0} \omega_1^{\rho_1} \cdots \omega_g^{\rho_g}$$

zu, so gehört zu jeder Einheit r eine und nur eine Einheitswurzel $\Omega(r)$, welche wir einen *Charakter* von r nennen wollen, denn durch r ist ja das Indexsystem (ρ, ρ_0, \dots) , also $\Omega(r)$ eindeutig bestimmt.

Here, again, to each system of values (k, k_0, \dots) and each unit r there clearly corresponds a character $\Omega^{(k)}(r)$.⁵¹ (Kronecker 1901, p. 445)

In contrast, say, to Hadamard, the index (k) is not arbitrary; rather, (k) is taken to range over the specific representing data. But, the notation and organization of the proof enables us to ignore the details of the representation where they are not needed. Kronecker noted immediately that $\Omega^{(k)}(r)$ depends only on the value of r modulo m , since any two values with the same residue have the same index. He also notes that we have $\Omega^{(k)}(rr') = \Omega^{(k)}(r)\Omega^{(k)}(r')$. The choice of representation has the further nice property that $\Omega^{(k)}(r)\Omega^{(k')}(r) = \Omega^{(k+k')}(r)$ where $(k+k')$ denotes the result of adding the elements of the tuples (k) and (k') , modulo the cardinality of the associated cyclic groups. Kronecker presented the first orthogonality principle:

$$\sum_{(r)} \Omega^{(0)}(r) = \varphi(m),$$

where r ranges over a system of residues of the units modulo m and (0) is the index of the trivial character, and

$$\sum_{(r)} \Omega^{(k)}(r) = 0$$

for the remaining characters. Kronecker's proof requires unfolding the notation and calculating, but thereafter, the fact can be recalled and used in the above form. Similarly, he expressed the second orthogonality relation by the equations

$$\sum_{(k)} \Omega^{(k)}(r_0) = \varphi(m),$$

when r_0 is congruent to 1 modulo m , and

$$\sum_{(k)} \Omega^{(k)}(r) = 0$$

otherwise. He also expressed the dependence of a Dirichlet series on a character in terms of a dependence on (k) , by writing

$$L^{(k)}(z) = \sum_{n=1}^{\infty} \frac{\Omega^{(k)}(n)}{n^z} = \prod_p \frac{1}{1 - \frac{\Omega^{(k)}(p)}{p^z}}.$$

⁵¹ "Wenn kein Mißverständnis zu befürchten ist, wollen wir im folgenden das zu Grunde gelegte Exponentensystem (k, k_0, k_1, \dots) kurz durch (k) und den zugehörigen Charakter einfacher durch

$$\Omega^{(k)}(r)$$

bezeichnen. Auch hier entspricht für ein festes Wertsystem (k, k_0, \dots) jeder Einheit offenbar r ein Character $\Omega^{(k)}(r)$."

Kronecker’s presentation provides us with an important lesson. The exposition is clearly designed to bring the central ideas to the fore and highlight relevant information, but succeeds in doing so while providing explicit representations and algorithms throughout. This shows that although the moves toward abstraction that we have documented in this section can sometimes *make it possible* to suppress or even ignore algorithmic information, they do not *require* one to do so. In other words, the conceptual reorganization opens the door to the use of other means of describing mathematical objects and operations on them, means that can supplant algorithmic aspects. The question as to whether such means are permissible, meaningful, and appropriate to mathematics lay at the core of twentieth-century foundational debates.

6.7 Landau

Born in 1877, Edmund Landau received his doctorate at the University of Berlin in 1899, having studied number theory under Frobenius. He completed a *Habilitation* thesis on Dirichlet series in 1901. Landau later presented proofs of Dirichlet’s theorem in two textbooks, his 1909 *Handbuch der Lehre von der Verteilung der Primzahlen* (Landau 1909) and his 1927 *Vorlesungen über Zahlentheorie* (Landau 1927). We shall here focus on his presentation in the 1909 work, since the later presentation is already essentially what we have portrayed as “contemporary” in Sect. 3.

Landau began by introducing the characters via the construction in terms of primitive roots and roots of unity. However, the notation that he used to denote them changed over the course of the book. Initially, he used a notation that is quite similar to Dirichlet’s notation for L -functions, writing

$$\chi(a_1, a_2, \dots, a_r, a, b)(n),$$

where the index system $a_1, a_2, \dots, a_r, a, b$ serves to distinguish the characters just as Dirichlet’s a, b, c, c', \dots distinguished his L -functions. However, after proving that there are $\varphi(k)$ such characters, he simplified his notation to

$$\chi_1(n), \chi_2(n), \dots, \chi_{\varphi(k)}(n),$$

and to $\chi_x(n)$ in the general case, thus adopting notation similar to both de la Vallée-Poussin and Hadamard.

But whereas Landau constructed the characters in the same way as Dirichlet, Dedekind, Hadamard, and de la Vallée-Poussin, he also recognized that they are characterized by certain key *properties* and that it is only these properties that are needed in the proof. Indeed, after introducing the abbreviations for the characters, Landau wrote:

...I will prove four short and elegantly-worded theorems about them [the characters]. The reader may then quickly forget the rather complicated definition of these functions completely, and need only remember that the existence of

a system of h distinct functions which possesses the four properties has been proved.⁵² (Landau 1909, p. 404)

The four theorems that Landau was referring to are the following:

Theorem 1 *For any two positive numbers n and n' ,*

$$\chi(nn') = \chi(n)\chi(n').$$

This “law of multiplication” holds for each of the h functions...

Theorem 2 *For $n \equiv n' \pmod{k}$,*

$$\chi(n) = \chi(n')$$

...

Theorem 3 *When n runs through a complete residue system modulo k , for $x = 1$, i.e. for the principal character*

$$\sum_n \chi_x(n) = h,$$

however for $x = 2, \dots, h$, i.e. for all other characters

$$\sum_n \chi_x(n) = 0$$

...

Theorem 4 *When n is fixed and the sum*

$$\sum_{x=1}^h \chi_x(n)$$

extends over all h functions, then

$$\begin{aligned} \sum_{x=1}^h \chi_x(n) &= h \quad \text{for } n \equiv 1 \pmod{k}, \\ \sum_{x=1}^h \chi_x(n) &= 0 \quad \text{for } n \not\equiv 1 \pmod{k}, \end{aligned}$$

⁵² “...ich werde über sie vier Sätze mit sehr kurzem und elegantem Wortlaut beweisen. Alsdann darf der Leser bald die recht komplizierte Definition dieser Funktionen vollkommen vergessen und braucht sich nur zu merken, daß die Existenz eines Systems von h verschiedenen Funktionen bewiesen worden ist, welche die vier Eigenschaften besitzen.”

therefore for all $k - 1$ other residue classes modulo k .⁵³ (Landau 1909, pp. 401–408)

These are exactly the four “important relations” given by de la Vallée-Poussin, and the analogues of the four properties enumerated by Dirichlet in 1840. What is novel here is Landau’s explicit recognition that only these properties are used in the proof and that the specific construction is only needed to show the existence of a system of functions that satisfy them.

Like Hadamard, Landau did not go so far as to allow the characters themselves to index the sum in Theorem 4 above, relying on a natural number proxy. But, like de la Vallée-Poussin, he gave an extensional classification of the three types of characters, although he also gave an additional description of the real characters in intensional terms.

⁵³ “**Satz 1:** Es ist für zwei ganze positive Zahlen n, n'

$$\chi(nn') = \chi(n)\chi(n').$$

Von jeder der h Funktionen wird also dies “Multiplikationsgesetz” behauptet ...

Satz 2: Es ist für $n \equiv n' \pmod{k}$

$$\chi(n) = \chi(n')$$

...

Satz 3: Wenn n ein vollständiges Restsystem modulo k durchläuft, ist für $x = 1$, d.h. für den Hauptcharakter

$$\sum_n \chi_x(n) = h,$$

dagegen für $x = 2, \dots, h$, d.h. für alle übrigen Charaktere

$$\sum_n \chi_x(n) = 0$$

...

Satz 4: Wenn n festgehalten und die Summe

$$\sum_{x=1}^h \chi_x(n)$$

über alle h Funktionen erstreckt wird, so ist

$$\sum_{x=1}^h \chi_x(n) = h \quad \text{für } n \equiv 1 \pmod{k},$$

dagegen

$$\sum_{x=1}^h \chi_x(n) = 0 \quad \text{für } n \not\equiv 1 \pmod{k},$$

also für alle $k - 1$ übrigen Restklassen modulo k .”

Just as Landau's notation for characters changed over the course of the book, so, too, did his notation for L -functions. Initially, he denoted them in a manner similar to Hadamard's, writing $L_x(s) = \sum_{n=1}^{\infty} \frac{\chi_x(n)}{n^s}$. However, later on, after concluding his presentation of the proof of Dirichlet's theorem, Landau adopted a "more convenient" notation:

Now let

$$L_x(s) = \sum_{n=1}^{\infty} \frac{\chi_x(n)}{n^s}$$

be the function corresponding to the character $\chi(n) = \chi_x(n)$; it is now more convenient to include the character in the notation,

$$L(s, \chi),$$

and, only when there is no fear of misunderstanding, write

$$L(s)$$

for short, as before.⁵⁴ (Landau 1909, p. 482)

Landau did not explain *why* the modern notation is more convenient, but we will offer some suggestions in the next section.

Landau's 1909 presentation is interesting because it has a transitional feel. Although he gave an intensional construction of the characters, he not only provided a thorough enumeration of their properties, but went out of his way to emphasize that these properties are all that matters to the proof. This is borne out by the fact that the division of characters into three classes can be carried out extensionally, without reference to their construction. And although he initially introduced natural number indices for the characters and L -functions and summed over these indices, he eventually adopted the modern functional notation for L -functions.

⁵⁴ "Es sei nun

$$L_x(s) = \sum_{n=1}^{\infty} \frac{\chi_x(n)}{n^s}$$

die dem Charakter $\chi(n) = \chi_x(n)$ entsprechende Funktion; es ist jetzt bequemer, um die Charakter in die Bezeichnung aufzunehmen,

$$L(s, \chi)$$

zu schreiben, und nur, wenn kein Mißverständnis zu befürchten ist, wie früher kurz

$$L(s)."$$

When he wrote his 1927 textbook Landau 1927, Landau finally made the transition to a proof that is extremely close to our contemporary version. Indeed, he no longer defined the characters by describing how they are constructed, but, rather, defined them in terms of their characteristic properties. Moreover, he summed over the characters themselves, using expressions such as $\sum_{\chi} \chi(a)$. The only real difference between his proof and the one we presented in Sect. 3.4 is that he did not develop the general notion of a group-theoretic character, but, rather, defined them in terms of the *particular* group of units modulo k . This makes sense, given that the work is an elementary number theory textbook and characters are not needed for any other purpose. Roughly speaking, if we combine Landau’s 1927 presentation with Weber’s 1882 more general treatment of characters, the result is the presentation in Sect. 3.

7 Changes in mathematical method

In Sect. 4, we considered a number of ways in which characters are treated as bona fide objects in contemporary proofs of Dirichlet’s theorem. We noted that none of these features are present in Dirichlet’s original proof, for the simple reason that Dirichlet did not isolate or identify the notion of a character. Dedekind’s 1863 presentation of Dirichlet’s theorem did not use the term “character,” but he did introduce the notation, $\chi(n)$, for characters and identified the defining property of a homomorphism as their “characteristic property.” As noted in Sect. 6.2, after Weber’s paper of 1882 the general notion of a group character was in place, and all the subsequently published proofs of Dirichlet’s theorem use the terminology of characters.

But now, the set of characters modulo m can be defined extensionally, as the set of nonzero homomorphisms from $(\mathbb{Z}/m\mathbb{Z})^*$ to the complex numbers, or intensionally, as functions defined by certain algebraic expressions involving certain primitive elements modulo the prime powers occurring in the factorization of m , and certain complex roots of unity. Even though the two definitions give rise to the same set of characters, proofs can differ in the extent to which they rely on the specific representations or the abstract characterizing property. Dirichlet’s proof relied only on the symbolic representations. Somewhat surprisingly, both Dedekind’s and Hadamard’s division of the characters into the trivial, real, and complex cases was also described in terms of the characters’ representations, even though the distinction is naturally expressed in terms of the values they take. Kronecker and de la Vallée-Poussin provided both descriptions, and even though Kronecker made it clear that all operations and classifications can be carried out, algorithmically, in terms of the canonical representations, his careful choice of notation and organization made the extensional properties salient. By 1927, Landau clearly favored the extensional characterization in his textbook.

We have also observed that modern notation like $\sum_{\chi} \chi(n)$ allows us to carry out summations over the finite set of characters modulo m , but that this notation was not used by Dirichlet’s early expositors. Dirichlet, Dedekind, and Kronecker all took summations to range over the tuples of integers representing the characters via an explicit algebraic definition, though Kronecker’s way of letting a variable (r) range over these tuples is more attractive than Dirichlet’s use of a, b, c, c', \dots . In 1882, Weber wrote his sums with ellipses. Curiously, Hadamard in 1896, and Landau in

1909, assigned arbitrary integer indices to the characters and took sums to range over those indices. Of the proofs of Dirichlet's Theorem 3.1 that we have considered, the only one that takes summation over characters at face value is that of de la Vallée-Poussin, who nonetheless adopted Dirichlet's 1840 notation S_χ to denote such sums. By 1927, however, Landau had adopted the modern notation.

Finally, we have emphasized the modern tendency to represent the dependence of an L -series on a character χ as a functional dependence, with the notation $L(s, \chi)$. Once again, de la Vallée-Poussin was the only nineteenth-century expositor of Dirichlet's theorem to do so, with the notation $Z(s, \chi)$. We saw that Landau made the transition in the middle of his 1909 book and that, in his 1927 textbook, he relied exclusively on the notation $L(s, \chi)$.

At issue in all these developments is whether characters could be treated in much the same way as natural numbers and real numbers, or whether characters are different sorts of objects, whose treatment has to be mediated by more "concrete" mathematical representations. We contemporary readers of the nineteenth-century literature are now so familiar with a modern perspective that it can be hard for us to appreciate the reasons it took so long for the community to adopt it. Let us begin our analysis of the history, then, by reflecting on the countervailing pressures.

We have had a lot to say about the notational and expressive conventions at play. The difficulty in settling on stable and useful conventions should not be minimized. For example, any mathematician writing a proof has to choose names for variables: Should one use m , n , and k to range over natural numbers, or x , y , and z ? One desiderata is to maintain continuity with the background literature, but other constraints come into play: For example, one reason to favor m , n , and k may be that x , y , and z are natural choices to range over other objects, like real or complex numbers, arising in the proof. It is significant, though not surprising, that Dedekind, 26 years after Dirichlet's proof was published, and de la Vallée-Poussin, fully 60 years later, both stuck with Dirichlet's choice of m and k for the first term and common difference in the arithmetic progression, respectively. Even today, q is often used to range over prime numbers in the definition of the L -series, and we still use the letter L exclusively, though most contemporary proofs use a and d in place of m and k .

In a similar way, one has to settle on choices of notation. This also requires thought, even in places where the norms that govern the notation are fairly clear. For example, assuming one divides the set C of characters modulo k into the set consisting of the trivial character, C_{triv} , the set of real characters, C_{real} , and the set of complex characters, $C_{complex}$, it is clear that a sum over C can be broken up accordingly:

$$\sum_{\chi \in C} F(\chi) = \sum_{\chi \in C_{triv}} F(\chi) + \sum_{\chi \in C_{real}} F(\chi) + \sum_{\chi \in C_{complex}} F(\chi).$$

But, even so, one has to settle on the notation to express this relationship, and as the history of Dirichlet's theorem shows, it can be a long time before a particular means of expression becomes standard. The modern notation emphasizes the parallels between summing over sets C of characters and summing over sets C of natural numbers, but it is by no means obvious that conflating the two is a good idea.

In Sect. 2.1, we briefly touched on Frege’s treatment of functions in the logical system of his *Begriffsschrift* (Frege 1879), developed in greater detail in *Grundgesetze der Arithmetik* (Frege 1893). Contemporary readers of the *Grundgesetze* may be struck by how many pages are devoted to explaining the syntax of the formal language. For those familiar with modern logic, Frege’s lengthy explanations seem fiddly and pedantic. But, for Frege, getting the grammatical rules worked out was much of the battle. The syntax of a language only seems trivial when you already know how to speak it.

But it would be a mistake to suggest that all the considerations that Dirichlet’s successors faced were “merely” notational. Means of expression often rely on substantial features of our understanding of the nature of that which is expressed. For example, consider the modern way of writing Dirichlet’s famous example of 1927:

$$f(x) = \begin{cases} a & \text{if } x \text{ is rational} \\ b & \text{otherwise.} \end{cases}$$

This bears superficial similarity to case-based definitions of number-theoretic functions, or piecewise definitions of analytic functions, that were familiar at the time. But it was novel to base a case distinction on the property of being rational, and indeed, the notation masks significant assumptions about what it means to define a function on the reals. Consider Frege’s brief account of the evolution of the function concept in 1891:

In the first place, the field of mathematical operations that serve for constructing functions has been extended. Besides addition, multiplication, exponentiation, and their converses, the various means of transition to the limit have been introduced—to be sure, people have not always been clearly aware that they were thus adopting something essentially new. People have gone further still, and have actually been obliged to resort to ordinary language, because the symbolic language of Analysis failed; e.g. when they were speaking of a function whose value is 1 for rational and 0 for irrational arguments.⁵⁵ (Frege 1891, p. 12)

Frege was not merely concerned to have a convenient notation to express the definition by cases. The definition of f above does that perfectly well, even today. The point is rather that the notation should come with clear rules of use. That is what Frege took to be lacking in the casual use of ordinary language and what he took his formal system to provide. Similar methodological concerns lie beneath the surface whenever we write $\lim_{n \rightarrow \infty} a_n$ to denote the limit of a sequence of real numbers, or $I + J$ to denote the sum of two ideals in a ring of algebraic integers. There is nothing tame about the infinitary operations that underlie the notation.

⁵⁵ “Erstens nämlich ist der Kreis der Rechnungsarten erweitert worden, die zur Bildung einer Funktion beitragen. Zu der Addition, Multiplikation, Potenzierung und deren Umkehrungen sind die verschiedenen Arten des Grenzüberganges hinzugekommen, ohne daß man allerdings immer ein klares Bewußtsein von dem wesentlich Neuen hatte, das damit aufgenommen werde. Man ist weiter gegangen und sogar genötigt worden, zu der Wortsprache seine Zuflucht zu nehmen, da die Zeichensprache der Analysis versagte, wenn z.B. von einer Funktion die Rede war, deren Wert für rationale Argumente 1, für irrationale 0 ist.”

In contrast, foundational questions regarding the use of characters may seem mild. After all, it is easy to represent the characters modulo a positive integer k , and any character is determined by the values it takes on the finitely many residues modulo k . Nonetheless, the use and treatment of characters in proofs of Dirichlet's theorem bears upon central questions regarding the use and treatment of functions more generally, specifically regarding the relationship between a function and its various representations.

Let us think of Dirichlet L -functions $L(s, \chi)$ in quasi-computational terms. Such a function should take, as input, a real (or complex) value s and a character χ and return a complex number. Let us set aside the (important) question as to what it means to take a real or complex number as input, or return one as output. What does it mean to accept a character as an input? Should one think of the character as being "presented" to the function as an infinite set of input–output pairs? Or as the list of finite values on the residues modulo k ? (In that case, does the same conception work for function arguments that are not finitely determined?) Should one think, rather, of χ as being some sort of procedure, or subroutine? If so, what sorts of procedures and subroutines are allowed in the definition of a functional with argument χ ? Perhaps, instead, we should identify χ , as Dirichlet did, with its representation in terms of an expression involving certain roots of unity. But recall that those representations relied on choices of primitive elements in the representation of k , although it turns out that the value of the L -function does not vary with different choices of the primitive elements. On our conception, does L somehow "depend" on the choices of primitive elements?

These issues arise not only with respect to functional notation, but also with respect to statements involving quantifiers over characters. For example, the proof of the second orthogonality lemma requires the following fact:

If n is relatively prime to k , there is a character χ such that $\chi(n) \neq 1$.

If we prove this as a separate lemma, we can then invoke the lemma for a given k to "obtain" a χ with the relevant property. But what exactly have we "obtained"? A table of values? A procedure? A representation? If the lemma constructs a χ via a representation of one sort, but our proof of the second orthogonality lemma relies on a different representation of characters, is it legitimate to apply the lemma in that context?

The modern theory of computability and the semantics of programming languages offers various ways of thinking about computer programs which take functions as arguments. The issues are subtle and complex. In contrast, modern mathematics followed a route whereby these subtleties are for the most part set aside. In particular, they are deemed incidental to the proof of Dirichlet's theorem. Roughly speaking, to make sense of a functional $F(f)$ with function argument f , set theory identifies f "canonically" with its extension, a set of input–output pairs, without concern as to how f is represented or how (and whether) one can "compute" $F(f)$. Set theory then imposes on mathematical language the restriction that the definition of such a functional F can only depend on the extension of f . (Or, put differently: modern mathematical conventions evolved to ensure the latter fact, and axiomatic set theory was designed to model and explain those conventions.)

In the nineteenth century, the answers to the questions raised above were not at all obvious. Indeed, they are still debated among logicians and foundationally minded mathematicians today. Even for those inclined to dismiss those questions as irrelevant to the proof of Dirichlet’s theorem, it would not have been immediately clear as to whether they really could be dismissed, and, if so, how that should be done. What may seem to be “merely notational” developments in the presentation of Dirichlet’s theorem were part and parcel of the broader mathematical community’s attempt to fashion an understanding of functions as objects that would better support the mathematics of the time.

Setting aside issues of meaning, there may also be concerns about *correctness*. Any choice of notation that draws on an analogy between different domains presupposes that the analogy is appropriate, which is to say that sufficiently many properties carry over and that there are sufficient safeguards to bar the ones that do not. For example, a sum of the form $\sum_{x \in S} t(x)$ generally makes sense when S is finite and the term $t(x)$ takes values on a domain with an addition that is associative and commutative, so that the order the one “runs through” the elements of S does not matter. Summing over infinite totalities is more subtle; it typically depends on having a notion of convergence for the domain in which $t(x)$ takes values and worrying about the order in which terms are summed. Now, choosing the notation $\sum_{\chi} t(\chi)$ may not be a good idea if it encourages users to transfer facts and properties between domains in an invalid way. This would explain why Dirichlet and de la Vallée-Poussin chose a special notation S_{χ} for such sums, as such a new symbol would not come with any unwanted baggage. Adopting the contemporary notation may have required some kind of assurance that the intended contexts would be sufficient to control for proper use.

In sum, mathematical conventions evolve, expand, and change. Any time a mathematician writes a line of text, he or she is situated in a tradition with implicit norms and conventions, and the line of text just written becomes part of that tradition. The fact that so many of these conventions are communicated implicitly does not make them any less important to mathematical understanding.

When a mathematician writes a proof, the intent is that the inferences contained therein will be deemed by his or her colleagues to be correct and justified. Where the inferences are instances of familiar patterns, one desires that they will be easily recognized as such, so that the reader’s effort is conserved for more substantial cognitive tasks. When the inferences rely on assumptions that may be considered dubious or push familiar patterns of reasoning into unexplored territory, there is greater concern not only as to whether the reasoning will be recognized as correct, but also as to whether it will be deemed *appropriate* to addressing the mathematical issues at hand. Thus, there are always strong pragmatic pressures to stick close to established convention, and one should not expect fundamental aspects of the language and methods of mathematics to change in novel ways, unless there are strong forces pushing for such change.

In the case of Dirichlet’s theorem, we have seen some of the ways in which Dirichlet’s successors chose to modify, or “improve,” his presentation. Now let us try to understand some of the perceived benefits. Kronecker’s proof was explicitly designed to fill in information that was absent from Dirichlet’s proof, in the form of explicit bounds on a quantity asserted to exist. Many of the other developments were explicitly

designed toward paving the way to useful generalizations. This was clearly Dedekind's intent, for example, in pointing out that the Euler product formula depends only on certain multiplicative properties of the terms occurring in the sum. Similarly, abstracting proofs of properties of characters on $(\mathbb{Z}/m\mathbb{Z})^*$ that rely on features specific to the integers modulo m paves the way to extending these properties to group characters more generally. Characters and their properties form the basis for representation theory, which has been an essential part of group theory since the turn of the twentieth century (Hawkins 1971; Mackey 1980). Authors such as Hadamard, de la Vallée-Poussin, and Landau were also interested in extending Dirichlet's methods to other kinds of Dirichlet series, which now play a core role in analytic number theory.

But it would be a mistake to attribute all the benefits of the expository innovations we have considered to increased generality. After all, these innovations play an equally important role in fostering a better understanding of Dirichlet's proof itself, by highlighting key features of the concepts and objects in question, motivating the steps of the proof, and reducing cognitive burden on the reader by minimizing the amount of information that needs to be kept in mind at each step along the way. It is true that these benefits often support generalization, but they do so in part by making our thinking *vis-à-vis* Dirichlet's proof itself more efficient.

For example, proofs of Dirichlet's theorem that rely on explicit representations of the characters require us to keep the details of the representation in mind. Recall Dirichlet's original representation of an arbitrary character:

$$\chi(n) = \theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \dots$$

Here the reader has to keep in mind that $\alpha, \beta, \gamma, \gamma'$, and so on are the indices of n with respect to primitive elements chosen in the decomposition of the abelian group $(\mathbb{Z}/m\mathbb{Z})^*$, and $\theta, \varphi, \omega, \omega'$, and so on are the corresponding roots of unity. This information has to be kept in mind throughout the proof, because the nature of the objects and the dependences of $\alpha, \beta, \gamma, \gamma'$ on n may play a role in licensing an inference or calculation. Moreover, recall that we obtain all the characters by expressing all the roots

$$\theta = \Theta^a, \varphi = \Phi^b, \omega = \Omega^c, \omega' = \Omega'^{c'}, \dots,$$

in terms of primitive roots of unity $\Theta, \Phi, \Omega, \Omega', \dots$, and letting a, b, \dots range over the appropriate exponents. Once again, this information has to be remembered throughout. Later proofs are easier to read simply because they do not require us to keep as much information in mind and highlight the relevant dependences when they are needed.

One way of achieving this is by reorganizing the proof in such a way that some of the relevant information is localized to particular facts and calculations. For example, even if one resorts to representations to prove the orthogonality relations, if this is the only place they are used, then they do not need to be ready to hand when these relations are invoked in a calculation later on, where other analytic expressions and their properties are the objects of focus. Thus, modularity reduces cognitive burden and makes it easier to keep track of the global structure of the argument, providing high-level outlines, or sketches, of the proof. Such restructuring paves the way to generality:

Isolating context-specific details in well-insulated modules means that one can adapt the proof by changing the modules while preserving their external interfaces. But, to repeat, generality is not the only benefit: The restructuring improves the readability of the original proof as well.

Even in situations where there is a lot of information in play at once, judicious notation and means of expression can make important relationships between the data more salient. For example, in Sect. 5, we saw that Dirichlet wrote

$$\begin{aligned} & \sum \frac{1}{q^{1+\rho}} + \frac{1}{2} \sum \frac{1}{q^{2+2\rho}} + \frac{1}{3} \sum \frac{1}{q^{3+3\rho}} + \dots \\ &= \frac{1}{K} \sum \Theta^{-\alpha_m a} \Phi^{-\beta_m b} \Omega^{-\gamma_m c} \Omega^{-\gamma'_m c'} \dots \log L_{a.b.c.c' \dots} \end{aligned}$$

where we would write

$$\sum_{\chi \in (\mathbb{Z}/q\mathbb{Z})^*} \overline{\chi(m)} \log L(s, \chi) = \varphi(q) \sum_{p \equiv m \pmod{q}} \frac{1}{q^s} + O(1).$$

The second presentation highlights the relationship between the series $L(s, \chi)$ and the characters; in particular, the value $L(s, \chi)$ is multiplied by the value of the conjugate character, $\overline{\chi}$, at m in each term of the sum on the left-hand side. Dirichlet’s notation obscures this relationship, since the logarithm of the appropriate L -series $L_{a.b.c.c' \dots}$ is multiplied by the character $\Theta^{-\alpha_m a} \Phi^{-\beta_m b} \Omega^{-\gamma_m c} \Omega^{-\gamma'_m c'} \dots$ and the reader must sift through this expression to notice that the exponents $-\alpha, -b, -c, -c', \dots$ correspond to the subscripts of the L -series.

The same problem applies to Hadamard’s method of using arbitrary indices for the characters: Given a character ψ_v , there is no natural name for the index of its conjugate $\overline{\psi}_v$. Of course, one can introduce such a notation, but that requires keeping track of the relationship between the two notations, that is, the conjugation of characters and the associated operation on indices. And grouping the characters into different classes requires grouping the indices into different classes, again yielding an uncomfortable duality. (At least Kronecker maintained a monist consistency, insisting that all operations on characters are operations on their representations. So when $\Omega^{(k)}$ is a character, Kronecker could write $\Omega^{(-k)}$ for its conjugate, since the conjugate character is obtained by negating the elements of the corresponding tuple.)

Similar considerations may explain why Landau changed his notation for L -functions in the middle of his 1909 work, from $L_x(s)$ to $L(s, \chi)$. In the paragraph following his notation change, Landau showed that the theory of L -functions can be reduced to L -functions that correspond to particular types of characters, called *proper characters*. Roughly, proper characters modulo k are those that cannot be obtained as a character modulo K where $K < k$. To show that the theory of L -series can be reduced in the appropriate way, he proved that if χ is an improper character modulo k and X is a corresponding proper character modulo K , we have (Landau 1909, pp. 482–483)

$$L(s, \chi) = \prod_{v=1}^c \left(1 - \frac{\varepsilon_v}{p_v^s} L_0(s, X) \right).$$

Here the ε_v are certain roots of unity, c is a natural number (which can be 0, depending on how many prime factors of k are contained in K), and the real part of s is assumed to be > 1 . If Landau had kept his original notation, the left-hand side of the above equation would be written as $L_x(s)$ where $\chi = \chi_x$. But how would we represent the right-hand side and, in particular, what index should we choose for X ? We should note that one obvious choice for an index, x' , would not be available, since Landau used this previously in connection with the conjugates of characters. In the text, Landau had identified a relationship between a character, χ , and the corresponding proper character, X ; having to translate this to a relationship between indices would only clutter the exposition. These issues are compounded when, in later sections, Landau wanted to obtain functional equations to relate $L(s, \chi)$ and $L(1-s, \bar{\chi})$ and in doing so referred to the distinction between proper and improper characters (see, e.g., Landau 1909 § 130). Using the old notation, we would need to keep track of three different subscripts to index the various characters.

Consider, finally, the issue of uniformity of notation. We have already discussed concerns associated with the notation \sum_{χ} for summation over characters. But there is an obvious benefit to using the same notation for summing over finite sets of characters and summing over finite sets of integers, namely that the two operations really do share common properties. Indeed, the transfer is immediate via the Hadamard trick of assigning an integer index to each character. The option of introducing a new symbol every time one needs to index sums by finite sets of objects is clearly untenable, as it would result in a confusing explosion of notations. It does not seem at all surprising that Dirichlet's 1841 notation S_{χ} was short-lived.

To sum up, then, we have identified a number of advantages to the rewritings of Dirichlet's proof considered in the last section:

- Essential properties of key objects (or expressions) were isolated, reducing the amount of information that someone reading the proof has to keep in mind at each step.
- The organization of proofs became more modular, with information localized to very specific parts of the proof.
- Expressions became more readable, since irrelevant details were suppressed, and the features that remained made dependences and relationships between terms more salient.
- Notation became more uniform, highlighting commonalities between different domains.

Changes like this often go hand in hand with attempts to generalize concepts and methods to other domains, since managing and controlling the volume of domain-specific detail tends to bring to the fore aspects of the proof that transcend these specifics. But they also contribute to a better understanding of Dirichlet's proof itself and make the proof easier to read and reproduce from memory.

Benefits such as these are often dismissed as merely "pragmatic" or "cognitive," but this downplays the fact that such considerations effectively shape and justify the norms

that guide our mathematical practice. The history and philosophy of mathematics need to take them seriously.

8 Conclusions

Our history of Dirichlet’s theorem has highlighted one mathematical development that encouraged a higher-order treatment of functions, but, of course, there were others. One important example is the development of functional analysis. In 1859, George Boole published *A Treatise on Differential Equations* (Boole 1859), in which he introduced the subject as a study of “variable quantities” subject to “known” relations between their differential coefficients. That work is noteworthy for its observation that differential operators can be viewed as algebraic expressions subject to certain laws. Integration was also viewed abstractly; in the expression $y = \int \varphi(x)dx + c$,

the symbol \int denotes a certain process of integration, the study of the various forms and conditions of which is, in a peculiar sense, the object of this part of the Integral Calculus. (Boole 1859, p. 2)

Modern functional analysis takes the view that operations on functions like differentiation and integration can themselves be viewed as functions, defined over spaces of other functions. In 1887, Vito Volterra published a seminal paper, “Sopra le funzioni che dipendono de altre funzioni” (“On functions that depend on other functions”) (Volterra 1887), which adopted such a viewpoint. In 1901, Hadamard published *Leçons sur le calcul des variations*, which helped establish the foundations of functional analysis and, in fact, introduced the term “functional.” In 1930, Volterra published an English translation of a series of lectures he had given at the University of Madrid in 1925 (Volterra 1930), which began with a lengthy discussion of the notion. After presenting particular examples, he said:

We shall therefore say that a quantity z is a *functional of the function* $x(t)$ in the interval (a, b) when it depends on all the values taken by $x(t)$ when t varies in the interval (a, b) ; or, alternatively, when a law is given by which to every function $x(t)$ defined within (a, b) (the independent variable within a certain function field) there can be made to correspond one and only one quantity z , perfectly determined, and we shall write

$$z = F \left[\left[x(t) \right] \right] .$$

This definition of a *functional* recalls especially the ordinary general definition of a *function* given by Dirichlet. (Volterra 1930, p. 4)

The chapter as a whole describes an outlook that is essentially the contemporary mathematical viewpoint, yet in a way that makes clear that the viewpoint was one that the mathematical community was still getting used to.⁵⁶

Twentieth-century foundational efforts served moreover to unify the function concept and provide a comprehensive framework to facilitate such higher-order treatment. In 1905, Borel, Baire, Lebesgue, and Hadamard engaged in their famous debate as to whether it makes sense to consider “arbitrary functions,” not given by any rule or law.⁵⁷ In 1914, Felix Hausdorff published his *Grundzüge der Mengenlehre* (Hausdorff 1914), dedicated to Georg Cantor. The book established a modern set-theoretic foundation and used it to support the development of point-set topology. After discussing the notion of an ordered pair, Hausdorff gave the concept of function its modern definition:

...we consider a set P of such pairs, satisfying the condition that each element of A occurs as the first element of one and only one pair p of P . In this way, each element a of A determines a unique element b , namely, the one to which it is connected in a pair $p = (a, b)$. We denote this element associated to a , which is determined by and dependent on a , by

$$b = f(a),$$

and we say that on A (i.e. for all elements of A) a *unique function* of a is defined. We view two such functions $f(a)$, $f'(a)$ as equal when and only when the corresponding sets of pairs P , P' are equal, that is, when, for *each* a , $f(a) = f'(a)$.⁵⁸ (Hausdorff 1914, p. 33)

Thus, in the early decades of the twentieth century, the modern view of a function took root. A function could map elements of any domain to any other; one could specify a function by specifying any determinate law; functions could serve as arguments to other functions; and functions could serve as elements of algebraic structures and geometric spaces.

Our study of the treatment of characters in number theory has focused on only one small part of the grand historical development that resulted in this way of thinking. Despite its narrow focus, the case study has illuminated some important factors that

⁵⁶ See also the Griffith Evans' helpful introduction to the 1959 Dover reprinting of Volterra's lectures (Volterra 1930).

⁵⁷ The “five letters” are translated as an appendix to Moore (1982), reproduced in Ewald (1996), volume 2, pp. 1077–1086.

⁵⁸ “...betrachten wir eine Menge P solcher Paare, und zwar von der Beschaffenheit, daß jedes Element a von A in einem und nur einem Paare p von P als erstes Element auftritt. Jedes Element a bestimmt auf diese Weise ein und nur ein Element b , nämlich dasjenige, mit dem es zu einem Paare $p = (a, b)$ verbunden auftritt; dieses durch a bestimmte, von a abhängige, dem a zugeordnete Element bezeichnen wir mit

$$b = f(a)$$

und sagen, daß hiermit in A (d. h. für alle Elemente von A) eine *eindeutige Funktion* von a definiert sei. Zwei solche Funktionen $f(a)$, $f'(a)$ sehen wir dann und nur dann als gleich an, wenn die zugehörigen Paarmengen P , P' gleich sind, wenn also, für *jedes* a , $f(a) = f'(a)$ ist.”

contributed to the modern view. We have emphasized the difficulties inherent in developing a coherent, rigorous way of treating functions as objects, as well as concerns as to how one could do so in a way that is appropriate to mathematics and preserves all the information that is essential to a rigorous argument. We have also explored some of the benefits associated with the methodological changes. Many of these accrue to the ability of the new language and notation to highlight central features and relationships of the objects in question, while suppressing details that make it harder to discern the high-level structure of a mathematical argument. We also emphasized the importance of representing mathematical concepts and objects in order to capture those features that are uniform across different domains of argumentation, so that these uniformities can be packaged and used in a modular way.

Finally, we have argued that understanding and evaluating the considerations that shape mathematical language and inferential practice is an important part of the history and philosophy of mathematics. Over time, mathematics evolves in such a way as to support the pursuit of distinctly mathematical goals. One of these is the goal of maintaining an inferential practice with clear rules and norms, one that allows its practitioners to carry out, communicate, and evaluate arguments that can become exceedingly long and complex. Another is the goal of promoting efficiency of thought, leveraging whatever features we can to extend our cognitive reach and transcend our cognitive limitations. The challenge remains that of developing appropriate ways of coming to terms with such “pragmatic” and “cognitive” considerations.⁵⁹

References

- Avigad, Jeremy. Methodology and metaphysics in the development of Dedekind’s theory of ideals. In (Ferreirós and Gray, 2006), 159–186.
- Avigad, Jeremy. 2006. Mathematical method and proof. *Synthese* 153: 105–159.
- Avigad, Jeremy. 2008. Understanding proofs. In (Mancosu, 2008), 317–353.
- Avigad, Jeremy. 2010. Understanding, formal verification, and the philosophy of mathematics. *Journal of the Indian Council of Philosophical Research* 27:161–197.
- Avigad, Jeremy, and Rebecca Morris. n.d. Character and object. In preparation.
- Beaney, Michael (ed.). 1997. *The Frege Reader*. Malden, MA: Blackwell Publishing.
- Boole, George. 1859. *A treatise on differential equations*. Cambridge: Macmillan and Co.
- Bottazzini, Umberto, and Jeremy Gray. 2013. *Hidden Harmony-Geometric Fantasies: The rise of complex function theory*. Berlin: Springer.
- Cantor, Georg. 1895. Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen* 46: 481–512.
- Cayley, Arthur. 1854. On the theory of groups, as depending on the symbolic equation $\theta^n = 1$. *Philosophical Magazine* 7:40–47. Reprinted in his *Collected Mathematical Papers*, volume 2, 123–130. Cambridge: Cambridge University Press. 1889.
- Chorlay, Renaud. n.d. Questions of generality as probes into 19th century analysis. To appear in K. Chemla, R. Chorlay, D. Rabouin, editors, *Handbook on Generality in Mathematics and the Sciences*.
- de la Vallée Poussin, Charles Jean. Démonstration simplifiée du théorème de Dirichlet sur la progression arithmétique. *Mémoires couronnés et autres mémoires publiés par L’Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique* 53:1895–1896.
- de la Vallée Poussin, Charles Jean. 1897. *Recherches analytiques sur la théorie des nombres premiers*. Brussels: Hayez.

⁵⁹ See Avigad (2006, 2008, 2010), and the essays in Mancosu (2008) for initial philosophical attempts in that direction, and Harel et al. (2008), Harel and Kaput (1991) for a pedagogical perspective.

- Dedekind, Richard. 1854. Über die Einführung neuer Funktionen in der Mathematik. Delivered as a *Habilitation* in Göttingen on June 30, 1854. Reprinted in (Dedekind, 1932), volume 3, chapter LX, 428–438, and translated by William Ewald as “On the introduction of new functions in mathematics” in (Ewald, 1996), volume 2, 754–762.
- Dedekind, Richard. 1888. *Was sind und was sollen die Zahlen?* Braunschweig: F. Vieweg & Sohn. A later edition reprinted in (Dedekind, 1932), Volume 3, Chapter LI, 335–391. The second edition, with a new preface, was published in 1893, and is translated by Wooster Beman as “The nature and meaning of numbers” in *Essays on the Theory of Numbers*, Chicago: Open Court, 1901; reprinted by New York: Dover, 1963. The Beman translation is reprinted, with corrections by William Ewald, in (Ewald, 1996), volume 2, 787–833.
- Dedekind, Richard. 1894. Über die Theorie der ganzen algebraischen Zahlen. Supplement XI to the fourth edition of (Dirichlet, 1863), 434–657. Reprinted in (Dedekind, 1932), volume 3, chapter 46, 1–222.
- Dedekind, Richard. 1932. *Gesammelte mathematische Werke*, volumes 1–3, edited by Robert Fricke, Emmy Noether and Øystein Ore, Braunschweig: F. Vieweg & Sohn. Reprinted by Chelsea Publishing Co., New York, 1968.
- Dirichlet, Johann Peter Gustav Lejeune. 1829. Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données. *Journal für die reine und angewandte Mathematik*, 4:157–169. Reprinted in (Dirichlet, 1889), 117–132.
- Dirichlet, Johann Peter Gustav Lejeune. 1835. Ueber eine neue anwendung bestimmter integrale auf die summation endlicher oder unendlicher reihen. *Abhandlungen der königlich Preussischen Akademie der Wissenschaften*, 391–407. Reprinted in (Dirichlet, 1889), 237–256.
- Dirichlet, Johann Peter Gustav Lejeune. 1837. Beweis eines Satzes über die arithmetische Progression. *Bericht über die Verhandlungen der königlich Preussischen Akademie der Wissenschaften Berlin*. Reprinted in (Dirichlet, 1889), pages 307–312.
- Dirichlet, Johann Peter Gustav Lejeune. 1837. Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält. *Abhandlungen der königlich Preussischen Akademie der Wissenschaften* 45–81. Reprinted in (Dirichlet, 1889), 313–342. Translated by Ralf Stefan as “There are infinitely many prime numbers in all arithmetic progressions with first term and difference coprime”, arxiv:0808.1408.
- Dirichlet, Johann Peter Gustav Lejeune. 1840. Über eine Eigenschaft der quadratischen Formen. *Journal für die reine und angewandte Mathematik*, 21:98–100. Reprinted in (Dirichlet, 1889), 597–502.
- Dirichlet, Johann Peter Gustav Lejeune. 1841. Untersuchungen über die Theorie der complexen Zahlen. *Journal für die reine und angewandte Mathematik* 22:190–194. Reprinted in (Dirichlet, 1889), 503–508.
- Dirichlet, Johann Peter Gustav Lejeune. 1863. *Vorlesungen über Zahlentheorie*. Vieweg, Braunschweig. Edited by Richard Dedekind. Subsequent editions in 1871, 1879, 1894, with “supplements” by Richard Dedekind. Translated by John Stillwell, with introductory notes, as *Lectures on Number Theory*, American Mathematical Society, Providence, RI, 1999.
- Dirichlet, Johann Peter Gustav Lejeune. 1889. *Werke*, edited by Leopold Kronecker. Berlin: Georg Reimer.
- Edwards, Harold M. 1980. The genesis of ideal theory. *Archive for history of exact sciences* 23: 321–378.
- Edwards, Harold M. 1989. Kronecker’s views on the foundations of mathematics. In *The history of modern mathematics*, ed. D.E. Rowe, and J. McCleary, 67–77. San Diego: Academic Press.
- Edwards, Harold M. 2007. Kronecker’s fundamental theorem of general arithmetic. In *Episodes in the history of modern algebra (1800–1950)*, ed. Jeremy Gray, and Karen Parshall, 107–116. Providence, RI: American Mathematical Society.
- Edwards, Harold M. 2009. Kronecker’s algorithmic mathematics. *Mathematical Intelligencer* 31: 11–14.
- Eisenstein, Gotthold. 1850. Eine neue Gattung zahlentheoretischer Funktionen, weche von zwei Elementen abhängen und durch gewisse lineare Funktional-Gleichungen definirt werden. *Bericht über die zur Bekanntmachung geeigneten Verhandlungen der königlich Preussischen Akademie der Wissenschaften zu Berlin*, 36–42. Reprinted in Eisenstein’s *Mathematische Werke*, volume 2, pages 705–711, Chelsea Publishing Company, New York, 1989.
- Euler, Leonhard. 1748. *Introductio in Analysin Infinitorum, tomus primus. Lausannae*. Publications E101 and E102 in the Euler Archive.
- Euler, Leonhard. 1784. Speculationes circa quasdam insignes proprietates numerorum. *Acta Academiae Scientiarum Imperialis Petropolitinae* 4. Publication E564 in the Euler Archive. Translated by Jordan Bell as “Speculations on some characteristic properties of numbers”, 2007, arXiv:0705.3929.
- Everest, Graham, and Thomas Ward. 2005. *An introduction to number theory*. London: Springer.

- Ewald, William (ed.). 1996. *From Kant to Hilbert: A source book in the foundations of mathematics, volumes 1 and 2*. Oxford: Oxford University Press.
- Ferreirós, José. 1999. *Labyrinth of thought: A history of set theory and its role in modern mathematics*. Basel: Birkhäuser.
- Ferreirós, José, and Gray Jeremy (eds.). 2006. *The architecture of modern mathematics*. Oxford: Oxford University Press.
- Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Louis Nebert. Translated by Michael Beaney in (Beaney, 1997), 84–129.
- Frege, Gottlob. 1891. *Function und Begriff*. Jena: Hermann Pohle. Reprinted in (Frege, 2002), and translated by Peter Geach as “Function and concept” in (Beaney, 1997). Page number references are to the original publication.
- Frege, Gottlob. 1893. *Grundgesetze der Arithmetik*. Jena: H. Pohle, volume 1, volume 2, 1903. Excerpts translated by Michael Beaney in (Beaney, 1997), 194–223 and 258–289.
- Frege, Gottlob. 2002. *Funktion-Begriff-Bedeutung*, edited by Mark Textor. Göttingen: Vandenhoeck and Ruprecht.
- Gauss, Carl Friedrich. 1801. *Disquisitiones Arithmeticae*. Leipzig: G. Fleischer. Reprinted in Gauss’ *Werke*, Königlich Gesellschaft der Wissenschaften, Göttingen, 1863. Translated with a preface by Arthur A. Clarke, New Haven: Yale University Press, 1966, and republished by New York: Springer, 1986.
- Hadamard, Jacques. 1896. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques. *Bulletin de la Société Mathématique de France* 24: 199–220.
- Harel, Guershon, Evan Fuller, and Jeffrey M. Rabin. 2008. Attention to meaning by algebra teachers. *Journal of Mathematical Behavior* 27: 116–127.
- Harel, Guershon, and James Kaput. 1991. The role of conceptual entities in building advanced mathematical concepts and their symbols. In *Advanced mathematical thinking*, ed. D. Tall, 82–94. Dordrecht: Kluwer.
- Hausdorff, Felix. 1914. *Grundzüge der Mengenlehre*. Leipzig: Veit and Company. Reprinted by Chelsea Publishing Company, New York, 1949.
- Hawkins, Thomas. 1971. The origins of the theory of group characters. *Archive for History of Exact Sciences* 7: 142–170.
- Klein, Felix. 1893. A comparative review of recent researches in geometry. *Bulletin of the American Mathematical Society* 2: 215–249.
- Kleiner, Israel. 1989. Evolution of the function concept: A brief survey. *The College Mathematical Journal* 20: 282–300.
- Kronecker, Leopold. 1870. Auseinandersetzung einiger eigenschaften der klassenzahl idealer komplexer zahlen. *Monatsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 881–882. Reproduced in (Kronecker, 1895–1930), volume I, pages 271–282.
- Kronecker, Leopold. 1882. *Grundzüge einer arithmetischen Theorie der algebraischen Grössen*. Rieme, Berlin. Also published in *Journal für reine und angewandte Mathematik*, volume 92, 1–122, and (Kronecker, 1895–1930), volume II, 237–387.
- Kronecker, Leopold. 1887. Ein Fundamentalsatz der allgemeinen Arithmetik. *Journal für die reine und angewandte Mathematik* 100:490–510. Reprinted in (Kronecker, 1895–1930), volume IIIa, pages 209–240.
- Kronecker, Leopold. 1901. *Vorlesungen über Zahlentheorie*, edited by Kurt Hensel, B. G. Teubner, Leipzig. Republished by Springer, Berlin, 1978.
- Kronecker, Leopold. 1895–1930. *Werke*, edited by Kurt Hensel, volumes 1–5, B. G. Teubner, Leipzig. Reprinted by Chelsea Publishing Co., New York.
- Landau, Edmund. 1909. *Handbuch der Lehre von der Verteilung der Primzahlen*, volume 1. Leipzig: B. G. Teubner.
- Landau, Edmund. 1927. *Vorlesungen über Zahlentheorie*. S. Hirzel, Leipzig. Translated by Jacob E. Goodman, as *Elementary number theory*, Chelsea Publishing Company, New York.
- Legendre, Adrien-Marie. 1788. Recherches d’analyse indéterminée. *Histoire de l’Académie Royale des Sciences de Paris*, 465–559.
- Luzin, Nikolai. 1998. Functions. *The American Mathematical Monthly* 105:59–67 (Part I) and 105:263–270 (Part II).
- Mackey, George W. 1980. Harmonic analysis as the exploitation of symmetry: A historical survey. *Bulletin of the American Mathematical Society* 3: 543–698.
- Mancosu, Paolo (ed.). 2008. *The philosophy of mathematical practice*. Oxford: Oxford University Press.

- Monna, A.F. 1972. The concept of function in the 19th and 20th centuries, in particular with regard to the discussions between Baire, Borel and Lebesgue. *Archive for History of Exact Sciences* 9: 57–84.
- Moore, Gregory H. 1982. *Zermelo's axiom of choice: Its origins, development, and influence*. New York: Springer.
- Morris, Rebecca. 2011. *Character and object*. Master's thesis. Carnegie Mellon University, Pittsburgh
- Pengelley, David. 2005. Arthur Cayley and the first paper on group theory. In *From Calculus to Computers: Using the Last 200 Years of Mathematics History in the Classroom*, ed. Amy Shell-Gellasch, and Dick Jardine, 3–8. Washington, D.C.: Mathematics Association of America.
- Stein, Howard. 1988. Logos, logic, and logistiké. In *History and Philosophy of Modern Mathematics*, ed. William Aspray, and Philip Kitcher, 238–259. Minneapolis, MN: University of Minnesota Press.
- Tappenden, Jamie. 2006. The Riemannian background to Frege's philosophy. In (Ferreirós and Gray, 2006), 97–132.
- Troelstra, A.S., Dirk van Dalen. 1988. *Constructivism in mathematics: An introduction*, volumes 1 and 2. Amsterdam: North-Holland.
- Volterra, Vito. 1887. Sopra le funzioni che dipendono de altre funzioni. *Rend. R. Accademia dei Lincei* 2:97–105, 141–146 and 153–158.
- Volterra, Vito. 1930. *Theory of Functionals and of Integral and Integro-differential Equations*. Blackie and Son, Ltd., London and Glasgow. Translated by M. Long and edited by Luigi Fantappiè. Reprinted with a preface by G. C. Evans, a biography of Vito Volterra and a bibliography of his published works by E. Whittaker, New York: Dover, 1959.
- Weber, Heinrich. 1882. Beweis des Satzes, dass jede eigentlich primitive quadratische Form unendlich viele Primzahlen darzustellen fähig ist. *Mathematische Annalen* 20: 301–329.
- Wilson, Mark. 2006. *Wandering significance: An essay on conceptual behavior*. Oxford: Oxford University Press.
- Wussing, Hans. 1984. *The Genesis of the Abstract Group Concept: A Contribution to the History of the Origin of Abstract Group Theory*, translated from the German by Abe Shenitzer and Hardy Grant. Cambridge, MA: MIT Press.
- Youschkevitch, A.P. 1976. The concept of function up to the middle of the 19th century. *Archive for History of Exact Sciences* 16: 37–85.