Character and Object

Becky Morris

December 28, 2011

1 Introduction

The concept of function permeates most branches of modern mathematics: from analysis, algebra and logic to probability and mathematical physics. However, the set-theoretic conception of function as a particular mathematical object (specifically, as a set of ordered pairs satisfying conditions of totality and single valuedness) was a long time in development. Indeed, historians generally cite the eighteenth century as the time during which the concept of function first made an explicit appearance. Moreover, from the time it was first introduced to the present day, the concept of function underwent significant development and changed considerably (see e.g. [Mon72], [You76], [Kle89], [Luz98a], and [Luz98b]). Most of the historical literature focuses primarily on the development of the function concept within mathematical analysis, but the concept also underwent significant changes within number theory. In this thesis, we will explore a facet of the development of the function concept in number theory by looking at the evolution of particular types of functions, and the associated philosophical consequences.

In particular, we shall examine the history of certain functions called Dirichlet characters and Dirichlet L-functions used in the proof of a result first established by Dirichlet in 1837 [Dir37a]. This proof was (re)presented and discussed by various other mathematicians in the nineteenth century, including Dedekind [DD99], de la Vallée Poussin [dlVP96], [dlVP97], Hadamard [Had96] and Landau [Lan09], [Lan27] and appears in many modern textbooks (for example Everest and Ward [EW05]). An analysis of these various presentations reveals a distinct change in the way in which the characters were conceived, and in particular, we will see that mathematicians gradually became more willing to treat the characters as mathematical *objects* in their own right. We shall refer to this as the process of *reification* of characters and will explain how it is to be understood more precisely throughout the rest of the present paper. Moreover, we will see that there are important consequences associated with the different methods of working with and conceiving of the characters.

We begin in section 2 with a review of the historical literature. In section 3, we introduce the terminology used in the modern presentation by Everest and Ward and give a sketch of the main steps involved. In section 4, we examine Dirichlet's original proof, while discussing the approaches of Dedekind, de la Vallée Poussin, Hadamard and Landau in section 5.

2 History of the Function Concept

2.1 Introduction of the concept of function

We shall here outline some of the main developments in the history of the function concept, mentioning events discussed by Monna [Mon72], Youschkevitch [You76], Kleiner [Kle89], and Luzin [Luz98a], [Luz98b]. In particular, we will follow Kleiner's presentation closely. The purpose of this chapter is not only to give the reader some important and interesting historical background, but also to sketch the development of the function concept *as presented in the secondary literature* and to highlight that the literature says very little about the use of functions in number theory. Thus this chapter will be significantly different in character to (most of) the remaining chapters of this thesis, which will not rely heavily on secondary sources.

Mathematics was without the function concept for much of its history. Indeed, both Monna and Youschkevitch insist that there was no explicit conception of function in ancient mathematics (for example [Mon72, 58] and [You76, 42-43]). While Youschkevitch suggests that the concept first emerged in the fourteenth century in the traditions of natural philosophy in Oxford and Paris (see [You76, 45-46]), Monna , Kleiner and Luzin highlight the late seventeenth century or eighteenth century as the period during which the concept first explicitly appeared (see [Mon72, 58], [Kle89, 282-283], [Luz98a, 59]). Certainly, the eighteenth century was a crucial time in the development of the function concept, for it was in 1748 that Euler's famous *Introductio in analysin infinitorum* was published. As Kleiner observes, this tome was the first to place the function concept explicitly at the center of analysis [Kle89, 284]. However, the definition of function that we find in this work is couched in terms of *analytic expressions*:

> A function of a variable quantity is an analytical expression composed in any manner from that variable quantity and numbers or constant quantities (Euler in [Kle89, 284]).

The key phrase *analytic expression* was not given an explicit definition, but, as Youschkevitch and Kleiner point out, Euler did attempt to illustrate what he intended it to mean with examples. In particular, expressions that are obtained via a number of different operations were to count as analytic expressions. The permissible operations included subtraction, multiplication, division, extraction of roots, exponentiation, use of the logarithmic and trigonometric functions, differentiation and integration [You76, 61-62], [Kle89, 284]. Moreover, Euler made the claim that any function can in fact be given as a *power series*, i.e. a series of the form $Az^{\alpha} + Bz^{\beta} + Cz^{\gamma} + Dz^{\delta} + ...$ where $\alpha, \beta, \gamma, \delta$ stand for "any numbers" (Euler's expression given in [You76, 62]). Thus, for Euler, a power series provided a canonical way of writing functions as analytic expressions.

2.2 The Vibrating String Controversy

Euler, however, later came to reformulate his definition of function and a well documented incident in mathematical physics, known as the *Vibrating String Controversy* was partly responsible for his reformulation. Before describing the controversy, however, we should note that mathematicians during this period made a particular assumption about the nature of analytic expressions: if two analytic expressions take the same values for the same arguments at all points on an interval, then they must do so everywhere [Kle89, 285], [Luz98a, 63]. This assumption captures the idea that, as Luzin puts it, that there is some "unity" to an analytic curve, i.e. something which links the different parts of the curve together (see [Luz98b, 263]).

The vibrating string controversy began with a paper published by the mathematician Jean d'Alembert, in which he investigated the motion of an elastic string whose ends were fixed and which was then plucked and released. We can think of the position of the string at a given time t as given by a particular function, $f_t(x)$, as shown in figure 1.

d'Alembert wanted to obtain a function that would describe the motion of the string for any time t. In his analysis, he demonstrated that it must satisfy a particular partial differential equation, now known as the *wave equation* [Kle89, 286]:

$$\frac{\partial^2 y}{\partial t^2} = a^2 \frac{\partial^2 y}{\partial x^2}.$$

By using certain "boundary conditions", i.e. further conditions that the function must meet, d'Alembert showed that the solution to the partial differential equation would be of the following form [Kle89, 286]:

$$y(x, t) = \frac{\phi(x + at) + \phi(x - at)}{2}.$$

However, Kleiner points out that, for d'Alembert, the function ϕ must have an additional property: it must be given by a *single* formula. That is to say, functions that were defined *piecewise*, with different expression governing their behavior on

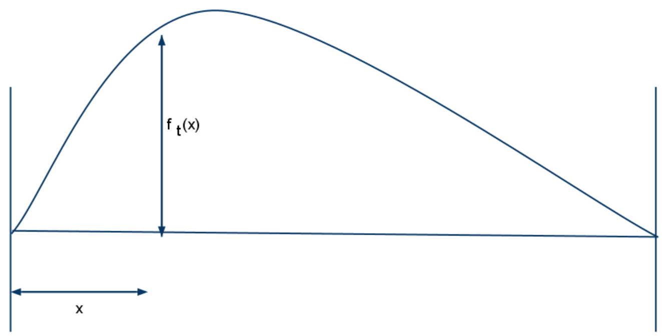

Figure 1: We can think of the position of the string at time t as given by a function $f_t(x)$.

different parts of their domain, were not admissible solutions for d'Alembert (see [You76, 65], [Kle89, 286]). Note that this means that the initial form of the string must also be described by a function that is given by a single formula. Indeed, if the initial position of the string is given by $f_0(x)$ then we must have $f_0(x) = \frac{\phi(x+a_0)+\phi(x-a_0)}{2} = \phi(x)$ (see [Kle89, 286]). As Luzin quotes d'Alembert:

> One cannot imagine a more general expression for a quantity y than that of supposing it to be a function of x and t; in which case the problem of the vibrating string has a solution only if the different forms of that string are contained in the same equation. (d'Alembert quoted in [Luz98a, 63]).

Euler, however, disagreed with d'Alembert's restriction upon the solution to the wave equation, maintaining instead that experimental [Kle89, 286-287] and geometrical [Luz98a, 63] considerations showed the solution to hold even in cases where the initial position of the string was not given by a single analytic expression. Thus, as Luzin describes, Euler raised the following challenge to d'Alembert: "If the obtained solution is to be regarded as deficient in those special cases when the form of the string cannot be contained by a single equation, what is one to mean by a solution in such cases?" (Euler in [Luz98a, 63]). In fact, Euler was willing to allow strings to have an initial position that was given by a function defined *piecewise* by different analytic expressions on different intervals, or even by a *freely drawn curve*[1] (see [Kle89, 287] and [Luz98a, 63]). And Euler and d'Alembert thought that it was not possible for the behavior of such functions to be described by a single analytic expression (see [Kle89, 287]). Why? Because it would clash with the assumption that any two analytic expressions that take the same values for the same arguments on an entire interval must take the same values for the same arguments everywhere. Indeed, this assumption tells us that any curve represented by a single analytic expression is determined by its behavior on any small interval (see [Kle89, 288]) and thus captures the intuition that a curve given by an analytic expression has a dependence between its parts. But surely, thought Euler and d'Alembert, a curve that was given by different analytic expressions on different intervals, or one that was drawn arbitrarily by hand could have no such dependency between its parts.

d'Alembert's response to Euler's account of what functions were admissible was to note that there were subtleties in the use of the differential equation that could not be overlooked. In particular, Luzin reports that as the solution to the wave equation must satisfy a particular differential equation, d'Alembert maintained that $\frac{\partial^2 y}{\partial x^2}$ must make sense and be finite (see [Luz98a, 63]). Having the initial shape of the string described by a function that is given by one or more different analytic expressions on different intervals of its domain or drawn freehand could allow problematic corners (see figure 2) and, according to d'Alembert, prevent the motion of the string (see [Luz98a, 63]).

Euler and d'Alembert were not the only two participants in this particular debate, however. Kleiner and Luzin describe the contribution of Daniel Bernoulli, who, in 1753, applied his knowledge of acoustics to the problem of determining the motion of a vibrating string (see [Kle89, 287] and [Luz98a, 63]). In particular, Bernoulli argued that the solution should be given by an infinite series of sines and cosines of multiple angles, i.e

$$y(x,t) = \alpha sin(\frac{\pi x}{l})cos(\frac{\pi at}{l}) + \beta sin(\frac{2\pi x}{l})cos(\frac{2\pi at}{l}) + \gamma sin(\frac{3\pi x}{l})cos(\frac{3\pi at}{l}) + ...$$

where the string has end points at $x = 0$ and $x = l$ (see [Kle89, 287], [Luz98a, 64]). Thus, given any starting position of a string with these endpoints, Bernoulli's solution meant that it could be represented as an infinite series of sines of multiple angles,

$$y(x,0) = f_0(x) = \alpha sin(\frac{\pi x}{l}) + \beta sin(\frac{2\pi x}{l}) + \gamma sin(\frac{3\pi x}{l})...$$

(see [Kle89, 287]).

[1] Euler had previously considered functions defined piecewise by analytic expressions, but did not appear to have considered functions drawn freely by hand (see e.g. [You76, 64,68]).

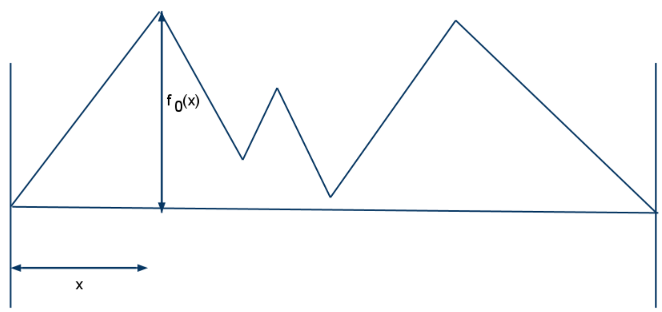

Figure 2: A string whose initial position is described by function $f_0(t)$ that has "corners".

Although Euler and d'Alembert had sharply disagreed about what sort of functions were admissible in the solution to the vibrating string problem, they were united in their view that Bernoulli was mistaken (see [You76, 66-67], [Kle89, 288] and [Luz98a, 64]). Their objection relied upon a consequence of the assumption concerning analytic expressions mentioned earlier: that a curve described by an analytic expression is determined in its entirety by its behavior on any small interval. Indeed, Euler reasoned that if we have some arbitrary function $f(x)$ then Bernoulli's work tells us that its curve can be given by an expression of the form $\alpha sin(\frac{\pi x}{l})+\beta sin(\frac{2\pi x}{l})+\gamma sin(\frac{3\pi x}{l})$.... But this, Euler argued, will be odd and periodic, and thus the "arbitrary" curve described by $f(x)$ must also be odd and periodic, which is absurd (see [Kle89, 288] and [Luz98a, 64]). d'Alembert upheld Euler's objection, and, Luzin reports, argued that not even all analytic periodic functions could be represented in the way Bernoulli was suggesting [Luz98a, 64].

Bernoulli's response, Luzin reports, was to appeal to the coefficients in the infinite series of sines and cosines. These could be chosen in a manner to allow the series to approximate an arbitrary curve as closely as was desired, he maintained (see [Luz98a, 64]). At the time, however, no method of calculating these coefficients was known, and Euler countered that if it was even possible, it would be prohibitively difficult to choose the coefficients in the manner Bernoulli had suggested (see [Luz98a, 64]). The debate between Euler, d'Alembert and Bernoulli failed to be settled and other mathematicians waded into the fray.

Although the controversy did not reach a resolution, it was partially responsible for a shift in Euler's conception of function, as mentioned earlier. Indeed, as Kleiner and Luzin have highlighted, Euler came to admit functions representing a curve drawn freely by hand into the initial conditions of the vibrating string problem. And such functions could not, from his perspective, be represented by an analytical expression. Thus his conception of function appears to have expanded, and as Kleiner emphasizes, this is reflected in his 1755 definition, where all mention of *analytic expressions* vanished:

> If, however, some quantities depend on others in such a way that if the latter are changed the former undergo changes themselves then the former quantities are called functions of the latter quantities. This is a very comprehensive notion and comprises in itself all the modes through which one quantity can be determined by others. If, therefore, x denotes a variable quantity then all the quantities which depend on x in any manner whatever or are determined by it are called its functions... [You76, 70] [Kle89, 288].

2.3 19th Century Developments

2.3.1 Fourier

The work of Jean Baptiste Joseph Fourier on heat transfer sparked new discussions on the concept of function in the nineteenth century. He had submitted his researches to the Paris Academy of Science in 1807, but it wasn't until 1822 that his work was finally published as *Théorie analytique de la chaleur* (see [Kle89, 289]). In this work, Fourier claimed to have established a result that conflicted with a number of eighteenth century ideas about the concept of function. In particular, his work denied the assumption that two analytic expressions that agree on an interval agree everywhere and the associated intuition that a curve given by an analytic expression possesses a certain unity. However, as Kleiner notes, Fourier's work itself was controversial, not just for challenging previously held ideas about functions, but because his arguments were not completely rigorous (see [Kle89, 289]). Let us now come to consider his work and its consequences in further detail.

In stating his result, Fourier elucidated the notion of function that he was working with as follows:

> In general, the function $f(x)$ represents a succession of values or ordinates each of which is arbitrary. An infinity of values being given to the abscissa x, there are an equal number of ordinates $f(x)$. All have actual numerical values, either positive or negative or null (Fourier in [Kle89, 289]).

He further clarified:

> We do not suppose these ordinates to be subject to a common law; they succeed each other in any manner whatever, and each of them is given as if it were a single quantity (Fourier in [Kle89, 289]).

With this notion of function, Kleiner presents Fourier's central result as follows [Kle89, 289]):

Theorem 2.1. *If $f(x)$ is a function defined on an interval $(-l, l)$, then*

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}[a_n cos(\frac{n\pi x}{l}) + b_n sin(\frac{n\pi x}{l})]$$

where

$$a_n = \int_{-l}^{l} f(t)cos(\frac{n\pi t}{l})dt \text{ and } b_n = \int_{-l}^{l} f(t)sin(\frac{n\pi t}{l})dt$$

We can easily see how Fourier's results violated the assumptions held by mathematicians working in the previous century. As Luzin sums up the consequences of his work,

> there is no organic connection between different parts of the same straight line or between different arcs of the same circle, since Fourier's discovery showed that one can subsume under a single analytic formula, a single equation, a continuous curve consisting of segments of different straight lines or arcs of different circles [Luz98a, 67].

2.3.2 Dirichlet

As Fourier's work was not of an appropriate standard of rigor, however, it was up to other mathematicians to refine and precisely formulate his ideas and, along the way, clarify the concept of function. Indeed, Fourier had claimed his result was true of *all* functions, and in order to make his work precise, it was necessary to know what was meant by "all functions".

Lejeune Dirichlet was one of the first mathematicians who made Fourier's work suitably rigorous. Indeed, in 1829, he published his *Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données* [Dir29]. In this paper, he considered a function defined on the interval $[-\pi, \pi]$, which he denoted by $\phi(x)$, and its associated Fourier series[2]:

$$\frac{a_0}{2\pi} + \frac{\sum_{n=1}^{\infty}[a_n cos(nx) + b_n sin(nx)]}{\pi}$$

[2]This is an alternative form of Fourier series to that presented in Kleiner. The difference between the two forms is due to the different intervals the function is defined upon.

where

$$a_n = \int_{-\pi}^{\pi} \phi(t)cos(nt)dt \text{ and } b_n = \int_{-\pi}^{\pi} \phi(t)sin(nt)dt$$

After much rigorous analysis, Dirichlet arrived at the following theorem ([Dir29, 131][3]):

Theorem 2.2. *...if the function $\phi(x)$, all of whose values are supposed finite and determined, presents only a finite number of discontinuities between the limits $-\pi$ and π, and if moreover it has only a determinate number of maxima and minima between the same limits, the series 7 [i.e. the Fourier series above], whose coefficients are the definite integrals dependent on the function $\phi(x)$, is convergent and has a value generally expressed by*

$$\frac{\phi(x+\varepsilon)+\phi(x-\varepsilon)}{2}$$

where ε designates a number infinitely small

Dirichlet's investigations did not end here, however. Indeed, he considered the case of a function that did not have finitely many discontinuities or finitely many maxima and minima (see [Dir29, 131-132]). In particular, he argued that it would not be possible for a function to be represented by its Fourier series on a particular interval if the set of discontinuities in that interval was dense, on the grounds that the notion of the integral of such a function would not make sense (see [Dir29, 131-132])[4]. To illustrate this with a concrete example, Dirichlet exhibited the now famous Dirichlet function, which is discontinuous everywhere:

$$f(x) = \begin{cases} c & \text{if } x \text{ is rational} \\ d & \text{if } x \text{ is irrational.} \end{cases}$$

This is significant, because, as Kleiner notes, it was the first example of a

[3]Les considérations précédentes prouvent d'une manière rigoureuse que, si la fonction $\phi(x)$, dont toutes les valeurs sont supposées finies et déterminées, ne présente qu'un nombre fini de solutions de continuité entre les limites $-\pi$ et π, et si en outre elle n'a qu'un nombre déterminé de maxima et de minima entre ces même limites, la série (7), dont les coefficients sont des intégrales définies dépendantes de la fonction $\phi(x)$, est convergente et a une valeur généeralement exprimée par:

$$\frac{\phi(x+\varepsilon)+\phi(x-\varepsilon)}{2},$$

où ϵ désigne un nombre infiniment petit.

[4]However, later conceptions of integration allowed for certain functions with a dense set of discontinuities to be integrated. See below.

function that had no analytic expression[5] and also could not be given graphically (see [Kle89, 292]).

And indeed, in a paper of 1837, *Über die Darstellung ganz willkürlicher Functionen durch sinus und cosinusreihen* [Dir37c], Dirichlet offered a definition of a (continuous) function [6]:

> One thinks of a and b as two fixed values and of x as a variable magnitude, which should assume all and only the values lying between a and b. Corresponding then to every x is a unique, finite y, and in such a way that, while x runs steadily through the interval from a to b, $y = f(x)$ also changes gradually. Then is y called a steady or continuous function. It is not necessary that y is dependent on the whole of this interval on some law of x, one does not even need to express the dependence through mathematical operations [Dir37c, 135].

Whilst some authors have taken this definition to apply to *all* functions, Youschkevitch notes that it explicitly applies only to *continuous* functions [You76, 78]. However, the fact that Dirichlet presented the Dirichlet function as a legitimate function suggests that he thought of his definition of continuous functions as applying more generally. Indeed, as Kleiner notes, the Dirichlet function "illustrated the concept of function as an arbitrary pairing" [Kle89, 290]. Thus, the conception of a function defined on an interval $[a, b]$ as an arbitrary correspondence is often referred to as the "Dirichlet definition" (see e.g. [Luz98b, 264]).

2.4 After the "Dirichlet definition"

In the years after Dirichlet exhibited the function that bears his name, more and more exotic functions appeared. In particular, Kleiner points to Riemann and Weierstrass. Riemann in his *Habilitationsschrift* [Rie67], published in 1867, extended the concept of integration and exhibited a function that had a dense set of discontinuities but which could nonetheless be integrated according to his extended conception (see e.g. [Kle89, 293]). And in 1872, Weierstrass exhibited a function

[5]This translation is my own, as are all others which are not otherwise cited. The original text will be included as a footnote for all my translations. It had no analytic expression at the time, but it was later discovered it could be written as $f(x) = (c - d)\lim_{n\to\infty}\lim_{m\to\infty}(\cos(n!\pi x))^{2m} + d$ [Kle89, 295].

[6]Man denke sich unter a und b zwei feste Wethe und unter x eine veränderliche Grösse, welche nach und nach alle zwischen a und b liegenden Werthe annehmen soll. Entspricht nun jedem x ein einziges, endliches y, und zware so, dass, währen x das Intervall von a bis b stetig durchläuft, $y = f(x)$ sich ebenfalls allmählich verändert, so heisst y eine stetige oder continuirliche. Es ist dabei gar nicht nötig, das y in diesem ganzen Intervalle nach demselben Gesetze von x abhängig sei, ja man braucht nicht einmal an eine durch mathematische Operationen ausdrückbare Abbhängigkeit zu denken.

that was continuous everywhere but differentiable *nowhere*, thus refuting claims made in numerous textbooks of the time (see e.g. [Kle89, 293]).

Famously, however, some mathematicians reacted strongly and negatively to such developments. In particular, Kleiner quotes Poincaré as follows:

> Logic sometimes makes monsters. For half a century we have seen a mass of bizarre functions which appear to be forced to resemble as little as possible honest functions which serve some purpose...In former times when one invented a new function it was for a practical purpose; today one invents them purposely to show up defects in the reasoning of our fathers and one will deduce from them only that [Kle89, 294].

Moreover, in the early twentieth century, a number of high profile mathematicians discussed the acceptability of the "Dirichlet definition" of function. In particular, Baire, Borel and Lebesgue[7] maintained that in order to legitimately define a function, an explicit *law of correspondence* must be included (see [Kle89, 296]). Borel imposed this requirement out of a concern for the communicability of mathematics: without an explicit law, how can two mathematicians know if they are discussing the same or different functions? Jacques Hadamard, however, vehemently resisted such restrictions:

> the requirement of a law that determined a function strongly resembles the requirement of an analytic expression for that function, and that this is a throwback to the eighteenth century [Kle89, 297].

Despite yet another controversy, the history of the function concept continued to develop fruitfully. Indeed, as Kleiner mentions, there were a number of more modern extensions, including the introduction of L_2 functions, distributions and the development of category theory (see [Kle89, 297-299]). In particular, in connection with the last of these three developments, Kleiner mentions the expansion of the permissible domain and codomain for mappings and refers to Dedekind. Dedekind's definition of mapping[8] in his 1888 essay *Was sind und was sollen die Zahlen?* [Ded88] is as follows:

> By a mapping of a system S a law is understood, in accordance with which to each determinate element s of S there is associated a determinate object, which is called the image of s and is denoted by $\phi(s)$;

[7]Baire and Lebesgue had been involved in a project to clarify the notions of "function" and "analytic expression". Baire had made precise a notion of "analytic expression", resulting in the Baire Classification scheme, whilst Lebesgue proved that each Baire Class was non-empty, and that there were functions that were not members of any Baire classes [Kle89, 295-296].

[8]The German word Dedekind uses is "Abbildung"

> we say too, that $\phi(s)$ corresponds to the element s, that $\phi(s)$ is caused or generated by the mapping ϕ out of s, that s is transformed by the mapping ϕ into $\phi(s)$. (Dedekind quoted in [Kle89, 299]).

Thus with Dedekind, we see the expansion of the concept of function to allow for functions defined not just on the real or complex numbers, but in fact on *arbitrary sets* [Kle89, 299]. Another mathematician associated with such an expansion is Volterra: for Volterra introduced functionals (functions that take a function as an argument and are real or complex values) [Kle89, 299]. And this is a notion of function that looks similar to the familiar, modern definition.

2.5 Summary

The historical development of the function concept is thus rich, detailed and complex. However, the above survey of the historical accounts found in Mona, Youschkevitch, Kleiner, and Luzin is indeed focused primarily on analysis. Kleiner's discussions briefly mention some of the more modern developments, and in particular, the expansion of the concept of function to include functions defined on sets whose arguments are not (subsets of) the real or complex numbers.

However, there is very little discussion of the development of functions within areas of mathematics outside of analysis. Moreover, there is also little mention about changes in the permissible and impermissible ways of treating them. This is quite understandable, however. The function concept itself arose primarily out of analysis, and as we have seen, the changes to the way in which functions were defined and what properties they were assumed to satisfy changed considerably even just within this domain, and it is important to understand them. However, changes in the conception of function also occurred within number theory during the nineteenth century, and they reflected a change in attitude concerning the nature of functions. In particular, there was a change in the way that mathematicians were willing to work with certain functions, e.g. from not permitting them to appear in the range of a bound variable, to allowing them to do so, and, we will argue, that changes like this indicate that the functions were being treated more like objects, and on a par with paradigmatic mathematical objects like the natural numbers. We maintain that understanding changes such as these, and the forces that drove them, are also important to understanding the modern conception of function.

3 Primes in Arithmetic Progressions

In the following sections, we will supplement the historical literature with considerations arising from a particular case study in number theory. In particular, we will look at proofs and discussions of the same theorem, called “Dirichlet’s theorem”, by Dirichlet, Dedekind, de la Vallée Poussin and Hadamard in their original historical context. We will identify the various ways in which certain functions are used in each of the proofs, and analyze both what this reveals about the conception of function and the impact such usage has upon the proof. Thus, in preparation for this, in the current section we will give a modern presentation of the key concepts and ideas required to prove Dirichlet’s theorem.

Let us begin with a statement of the theorem that is the subject of our case study, Dirichlet’s theorem on primes in arithmetic progressions:

Theorem 3.1. *Suppose that a and q are co-prime natural numbers. Then the arithmetic progression a, $a+q$, $a+2q$, ..., $a+nq$, ... will contain infinitely many prime numbers. In other words, there are infinitely many primes p such that $p \equiv a \pmod{q}$.*

This is related to the theorem that there are infinitely many primes, and indeed Dirichlet himself cited Euler’s analytic work on the distribution of prime numbers [Eul48, chapter 15] in a paper announcing his proof of the above theorem (see [Dir37b, 309-310]). As Dirichlet indicated in that paper, there are strong analogies between Euler’s work on the number and distribution of the prime numbers and his own on arithmetic progressions. So before proceeding to examine the modern proof of Dirichlet’s theorem, we turn to consider a modern presentation of Euler’s work on the distribution of the primes in his *Introductio*.

3.1 Prelude on the Infinity of the Primes

Whilst Euclid had proved the infinity of the primes in approximately 300 BC, his proof did not provide a great deal of information about how they were distributed, an issue about which Euler and other eighteenth century mathematicians were very curious. Thus Euler, in chapter *XV* of his *Introductio*, proved a theorem which is expressed in modern terminology as follows:

Theorem 3.2. *The series $\sum_p \frac{1}{p}$ is divergent, where the sum is over all primes p*

This theorem not only tells us that there are infinitely many primes (since the sum is divergent, the terms of the series, which are the reciprocals of the prime numbers, must be infinite in number), it also tells us something about the density of the primes. In particular, since we know that the series $\sum_n \frac{1}{n^2}$ is convergent, it tells us that the prime numbers occur with greater frequency than squares of natural numbers. Thus there are “more” primes than there are square numbers (see e.g. [Cop04, 428]).

Let us now present a very brief outline of a modern version of Euler's proof following the presentation given in detail in Noah Snyder's senior thesis and lecture notes [Sny02]. We will not fill in the details, but only indicate the main steps needed to establish the result.

The proof centers around the following infinite series:

Definition 3.3.

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

where s is a real variable.[9]

The series $\zeta(s)$ converges uniformly on the interval $[a, \infty)$ where a is any number greater than 1.

First, we prove a very important identity about this series, which is called the Euler-product formula:

Theorem 3.4. *For $s > 1$:*

$$\sum_{n=1}^{\infty} n^{-s} = \prod_{p} (1 - \frac{1}{p^s})^{-1}$$

where the product is over all primes p.

Assuming this has been proved, we take logarithms of both sides of the Euler-Product formula and appeal to properties of the logarithm to obtain the following equation:

$$\log \sum_{n=1}^{\infty} n^{-s} = \sum_{p} -\log(1 - \frac{1}{p^s})$$

The next step is to use the Taylor series expansions for $\log(1-x)$, and to change the order of the summations. After we do this, we obtain the following equation:

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

[9]The Riemann zeta function is obtained via analytic continuation of

$$\sum_{n=1}^{\infty} n^{-\sigma}$$

where $\sigma = s + it$, s, t real variables.

Keep in mind that we want to show that $\sum_p \frac{1}{p}$ diverges, and notice that the first term on the right hand side of the above equation is $\sum_p \frac{1}{p^s}$. Thus we should consider what happens as s tends to 1 from above. When we take the limit, the second term of the right hand side of the above equation, $\sum_{n=2}^{\infty} \frac{1}{n} \sum_p \frac{1}{p^n s}$, becomes small and negligible, but the left hand side, $\log \sum_{n=1}^{\infty} n^{-s}$, becomes infinite. Thus, as we take the limit, $\sum_p \frac{1}{p^s}$, must also become infinite. Once we have established this, it follows straightforwardly that $\sum_p \frac{1}{p}$ diverges.

3.2 Modern sketch of Dirichlet's theorem

Just as the modern presentation of Euler's proof demonstrates the infinitude of the prime numbers by establishing the series $\sum_p \frac{1}{p}$ is divergent, so the modern proof of Dirichlet's theorem establishes that for a, q co-prime natural numbers there are infinitely many primes p such that $p \equiv a \pmod q$ by showing that the series $\sum_{p \equiv a \pmod q} \frac{1}{p}$ is divergent. However, the proof of Dirichlet's theorem is considerably more complex than the proof of the divergence of the reciprocals of the primes. First of all, we need to consider functions that are more general than the series $\sum_{n=1}^{\infty} n^{-s}$ considered in the previous proof, and which are called L-functions. Before we can define the L-functions, we first have to introduce two other types of functions: Groups characters and Dirichlet characters. The material in this chapter concerning group characters, Dirichlet characters, L-functions, and the outline of the modern presentation of Dirichlet's proof itself follows closely the presentation given in Everest and Ward [EW05, 207-224].

3.2.1 Group Characters

Recall that a group G is a set equipped with an operation $\cdot : G \times G \to G$ and a distinguished element 1_G satisfying the following conditions:

1. Associativity. For any f, g, h in G, $(f \cdot g) \cdot h = f \cdot (g \cdot h)$
2. Identity. For any f in G, $f \cdot 1_G = 1_G \cdot f = f$
3. Inverses. For any f in G there exists an element f' such that $f \cdot f' = f' \cdot f = 1_G$

We will be interested in a particular type of groups, finite abelian groups, since it is on these that group characters are defined. A finite abelian group is defined as follows:

Definition 3.5. *A finite abelian group is a group G with only finitely many elements and which satisfies the axiom of commutativity, i.e. for any f, g in G $f \cdot g = g \cdot f$.*

An example of a finite abelian group is the multiplicative group of units of $\mathbb{Z}/n\mathbb{Z}$, also called the multiplicative group of units modulo n and denoted by $U(\mathbb{Z}/nZ)$. It is defined as follows:

Definition 3.6. *Group of units modulo n: $U(\mathbb{Z}/n\mathbb{Z})$ is the set of congruence classes $1 \pmod{n}, \ldots, j \pmod{n}, \ldots, n-1 \pmod{n}$ such that j, n are coprime, with the operation of multiplication modulo n.*

We can then define a group character as follows:

Definition 3.7. *A character χ of a finite abelian group G is a group homomorphism from G to $(\mathbb{C}^*, \cdot)$, i.e. the multiplicative group $\mathbb{C}\backslash\{0\}$ where $\mathbb{C}$ denotes the set of complex numbers. In other words, χ is a function from G to $\mathbb{C}\backslash\{0\}$ that satisfies the condition: for any $g, h \in G$, $\chi(g \cdot h) = \chi(g) \cdot \chi(h)$.*

As an example of a group character, consider the *trivial character*. This is defined for any finite abelian group G as follows:

Definition 3.8. *The trivial character $\chi_0 : G \to \mathbb{C}^*$ is defined by $\chi(g) = 1$ for all $g \in G$.*

As a slightly more complicated example of group characters, consider $U(\mathbb{Z}/10\mathbb{Z})$. It has four group characters defined on it as shown below:

Table 1: The group characters on the group of units modulo 10 $U(\mathbb{Z}/10\mathbb{Z})$

	χ_0	χ_1	χ_2	χ_3
1	1	1	1	1
3	1	i	-1	$-i$
7	1	$-i$	-1	i
9	1	-1	1	-1

There are a number of facts about group characters which we should recognize in order to understand the proof of Dirichlet's theorem. The first two observations to make are that for any character χ of any finite abelian group G, $\chi(1_G) = 1$ and that for any $g \in G$, $\chi(g)$ is a root of unity. To see why $\chi(1_G) = 1$, observe that $\chi(1_G) = \chi(1_G \cdot 1_G) = \chi(1_G)\chi(1_G)$ and so $\chi(1_G) = \chi(1_G)^2$. Thus $\chi(1_G) = 1$ or 0, but $\chi(1_G) \neq 0$ since 0 is not an element of $\mathbb{C}^*$. To see that $\chi(g)$ is a root of unity, recall that for every element g of G there is a natural number n such that $g^n = 1_G$. Then observe that $\chi(g)^n = \chi(g^n) = \chi(1_G) = 1$.

Let us now consider some more substantial theorems about group characters. The first is that the characters of a finite abelian group G themselves form a group called $\widehat{G}$ under the operation defined by $(\chi \cdot \psi)(g) = \chi(g)\psi(g)$. In fact, we can say something stronger:

Theorem 3.9. *[G is isomorphic to $\widehat{G}$] Let G be a finite abelian group, and $\widehat{G}$ the group of characters on G. Then $G \cong \widehat{G}$, and, in particular, if $|G|$=n then $|\widehat{G}| = n$.*

The proof of this theorem relies on an important result from algebra that tells us we can decompose a finite abelian group into a direct product of finite cyclic groups, where a group C with operation $\cdot$ is called cyclic if there exists an element $c \in C$ such that $C = \{c^n : n \in \mathbb{N}\}$. More formally, the result is as follows (for more details, see e.g. [CFR11]):

Theorem 3.10. The Structure Theorem for Finite Abelian Groups. *Suppose that G is a finite abelian group with operation $\cdot$. Then G is isomorphic to a direct product of cyclic groups, i.e. $G \cong \prod_{j=1}^{k} C_{n_j}$, where C_{n_j} denotes a cyclic group of order n_j.*

With this in mind, let us now consider the proof of the previous theorem.

Proof. Let us first establish the result for a cyclic group and then use the Structure Theorem for Finite Abelian groups to extend it to the more general case.

Let C_m be a cyclic group of order m with generator c_m. Then $C_m = \{1, c_m, \ldots, c_m^{m-1}\}$, i.e. we can write each element of C_m as c_m^l for some $l \in \{0, 1, \ldots, m-1\}$. Now define m functions, which we will denote by χ_0, χ_1 , ..., χ_{m-1}, from C_m to $\mathbb{C}^*$ as follows: $\chi_j(c_m^l) = \exp(\frac{2\pi i j l}{m})$. The values of each of these m functions are then mth roots of unity and it is easily checked that the χ_j are characters. Thus there is a mapping from C_m to $\widehat{C_m}$ which maps c_m^j to χ_j. And it is injective, since distinct elements c_m^r and c_m^s where $r, s \in \{0, 1, \ldots, m-1\}$ map to χ_r and χ_s respectively, whose values for the generator c_m are distinct mth roots of unity.

To show that this mapping is also surjective, suppose that we have a character ψ of C_m. As c_m is a generator of C_m we have that $c_m^m = 1_{C_m}$. Thus as ψ is a character we know that $\psi(c_m)^m = \psi(c_m^m) = \psi(1_{C_m}) = 1$. Thus $\psi(c_m)$ is an mth root of unity, i.e. it is $\exp(\frac{2\pi i j}{m})$ for some $j \in \{0, 1, \ldots, m-1\}$. But $\psi(c_m^l)$ is then determined for all $l \in \{1, 2, \ldots, m-1\}$, and thus we see that ψ is χ_j as defined above. Thus the mapping from C_m to $\widehat{C_m}$ is a bijection. Moreover, it is easy to see that it respects the group operations, and thus the mapping is in fact an isomorphism.

To apply this to the more general case, let G be a finite abelian group. By the Structure Theorem for Finite Abelian Groups, $G \cong \prod_{j=1}^{k} C_{n_j}$ where each of the C_{n_j} are finite cyclic groups of order n_j. Let c_j be a generator for C_{n_j}. Then we can represent each element $g \in G$ as $c_1^{l_1} \ldots c_j^{l_j} \ldots c_k^{l_k}$ where each $l_j \in \{0, \ldots, j-1\}$. We can then define functions on G as follows: Let $\chi^{(j)}(c_1^{l_1} \cdots c_j^{l_j} \cdots c_k^{l_k}) = \exp(\frac{2\pi i l_j}{n_j})$ for $j \in \{1, \cdots, k\}$, so these functions ignore all but the jth factor in the representation of the group element. It is easily checked that the $\chi^{(j)}$ are characters. Each of the $\chi^{(j)}$ generates a cyclic group of characters of order n_j, i.e. it generates characters

$\chi_0^{(j)}, \dots \chi_l^{(j)}, \dots \chi_{n_j-1}^{(j)}$ where $\chi_l^{(j)} = (\chi^{(j)})^l$. Moreover, the intersection of each such group with the span of the other characters will contain only the trivial character. Thus we have an injection from the elements of the group G to $\widehat{G}$.

To show that the map is surjective, suppose that we have a character ψ on the group G. As each element of the group G can be written as a product of generators of cyclic groups in the decomposition of G, and as ψ satisfies $\psi(gh) = \psi(g)\psi(h)$ for all $g, h \in G$, ψ will be determined by its values for the generators c_j. As each c_j generates C_{n_j}, we must have $c_j^{n_j} = 1_{C_{n_j}} = 1_G$. Thus as before $\psi(c_j)^{n_j} = \psi(c_j^{n_j}) = \psi(1_{C_m}) = 1$ and so $\psi(c_j)$ must be an n_jth root of unity, say $\exp(\frac{2\pi i t}{n_j})$ where $t \in \{0, \dots, n_j - 1\}$. Thus $\psi(c_j)$ is $\prod_{j=1}^{k} \chi_t^{(j)}$. Finally, it is easy to see that the map between the elements of G and the elements of $\widehat{G}$ respects the group operations, and is thus an isomorphism. $\square$

Now that we know a little more about the structure of characters on a finite abelian group G, we can prove some relations that play an important role in the proof of Dirichlet's theorem. These relations are known as the *Orthogonality relations*. Before proceeding to state and prove them, however, we first need a lemma.

Lemma 3.11. *If G is a finite abelian group and $g \in G \backslash \{1_G\}$, then there is a character $\chi \in \widehat{G}$ such that $\chi(g) \neq 1$.*

Proof. As we saw in the proof of the theorem that G is isomorphic to $\widehat{G}$, any element $g \in G$ can be expressed as a product of powers of generators of the cyclic groups in its decomposition. So, if $g \neq 1_G$ then $g = \prod_{i=1}^{k} c_i^{l_i}$ with $l_i \in \{0, \dots, n_i - 1\}$ and at least one factor $c_j^{l_j}$ is not the identity so that $l_j \geq 1$. But then $\chi_1^{(j)}$ as defined in the previous proof is such that $\chi_1^{(j)}(g) = \chi_1^{(j)}(\prod_{i=1}^{k} c_i^{l_i}) = \exp(\frac{2\pi i l_j}{n_j}) \neq 1$. $\square$

With this lemma, we can now go on to state and prove the orthogonality relations.

Theorem 3.12. *The Orthogonality Relations: If G is a finite abelian group, then for any element g of G and any character ψ of $\widehat{G}$, we have the following:*

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

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

Proof. We prove the first equation first. If ψ is the trivial character χ_0, then certainly $\sum_{h \in G} \psi(h) = 1$, since $\psi(h)$ will take the value 1 for each element h in the group G. So suppose that ψ is not the trivial character. Then we can find an element h' in the group such that $\psi(h') \neq 1$. Then note that $\psi(h') \sum_{h\in G} \psi(h) = \sum_{h\in G} \psi(h')\psi(h) = \sum_{h\in G} \psi(h'h) = \sum_{h\in G} \psi(h)$ since $h'h$ runs over the elements of G as h does. Thus $(\psi(h') - 1) \sum_{h\in G} \psi(h) = 0$. But, as $\psi(h') \neq 1$, we must have that $\sum_{h\in G} \psi(h) = 0$.

We now consider the second equation. The case when $g = 1_G$ is trivial because then for all characters χ, $\chi(g) = 1$ and thus $\sum_{\chi \in \widehat{G}} \chi(g) = |G|$. So suppose that $g \neq 1$. Then, by the previous lemma, we can find a character χ' such that $\chi'(g) \neq 1$. We now argue along the same lines as for the previous equation: $\chi'(g) \sum_{\chi \in \widehat{G}} \chi(g) = \sum_{\chi \in \widehat{G}} (\chi' \cdot \chi)(g) = \sum_{\chi \in \widehat{G}} \chi(g)$ since $\chi' \cdot \chi$ runs over the elements of $\widehat{G}$ as χ does. Thus we have $\chi'(g) \sum_{\chi \in \widehat{G}} \chi(g) = \sum_{\chi \in \widehat{G}} \chi(g)$. But as we know that $\chi'(g) \neq 1$, we must have that $\sum_{\chi \in \widehat{G}} \chi(g) = 0$. □

The orthogonality relations can then be used to prove another important result:

Corollary 3.13. *For any $g, h \in G$ we have the following*[10]*:*

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

Proof. To prove this result, first recall that for any h, $\chi(h)$ is a root of unity. Then note that $\chi(h^{-1}) = \frac{1}{\chi(h)} = \overline{\chi(h)}$. We then obtain the following from the second orthogonality relation by replacing g with gh^{-1}:

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

□

3.2.2 Dirichlet characters

Now that we have studied the structure of group characters, we can use them to introduce number theoretic functions that will be used directly in the modern presentation of Dirichlet's theorem on primes in arithmetic progressions.

Dirichlet characters are an extension of group characters on the group of units modulo n. In particular, they extend the group characters so that instead of being

[10] $\overline{\chi(g)}$ denotes the complex conjugate of $\chi(g)$, i.e. if $\chi(g) = u + iv$ then $\overline{\chi(g)} = u - iv$, where u, v are real numbers.

defined for certain residue classes, they are defined for all natural numbers. The formal definition is as follows:

Definition 3.14. *Dirichlet characters: Let* $1 < q \in \mathbb{N}$ *and fix a character* χ *on the finite abelian group* $U(\mathbb{Z}/q\mathbb{Z})$*. Then we can define a number theoretic function called a Dirichlet character (modulo q) X corresponding to* χ *as follows:*

$$X(n) = \begin{cases} \chi(n \pmod q) & \textit{if n is coprime to q} \\ 0 & \textit{otherwise} \end{cases}$$

Definition 3.15. *The principal character: The principal character modulo q* X_0 *is the Dirichlet character that corresponds to the trivial character* χ_0 *on* $U(\mathbb{Z}/q\mathbb{Z})$*.*

Thus, as a concrete example, let's consider the Dirichlet character obtained from χ_3 in the table 1. It will be as follows:

$$X(n) = \begin{cases} 0 & \text{if } n \text{ is not coprime to } 10 \\ 1 & \text{if } n \equiv 1 \pmod{10} \\ -i & \text{if } n \equiv 3 \pmod{10} \\ i & \text{if } n \equiv 7 \pmod{10} \\ -1 & \text{if } n \equiv 9 \pmod{10} \end{cases}$$

Before coming to introduce L-functions, we should note one crucial property that the Dirichlet characters enjoy:

Lemma 3.16. *A Dirichlet character X is* completely multiplicative*, i.e.*

1. $X(1) = 1$
2. $X(mn) = X(m)X(n)$ *for all natural numbers m and n.*

Proof. Let χ be the group character on $G = U(\mathbb{Z}/q\mathbb{Z})$ corresponding to the Dirichlet character X. We have already seen that for the identity element $\chi(1_G) = 1$. Thus by definition of the Dirichlet character X, $X(1) = \chi(1_G) = 1$. Thus the first condition is satisfied. Now let m, n be natural numbers. Suppose first of all that at least one of m, n are not coprime to q. Then the product mn will also not be coprime to q. Thus $X(mn) = X(m)X(n) = 0$. Now suppose that both m and n are coprime to q. Then we note that $(m \bmod q)(n \bmod q) = (mn \bmod q)$. Thus as χ is a group character, $X(mn) = \chi(mn \pmod q) = \chi(m \pmod q)\chi(n \pmod q) = X(m)X(n)$.

□

We should note that there is a potentially confusing notational convention concerning the group and Dirichlet characters: mathematicians use the symbol 'χ' to refer to both the group character χ and the corresponding Dirichlet character X. As this is the standard notation, however, we shall also adopt it from now on.

3.2.3 *L*-functions

Now that we have introduced and briefly studied group characters and Dirichlet characters, we can introduce *L*-functions, which are a generalization of the series $\sum_{n=1}^{\infty} \frac{1}{p^s}$ in Euler's proof of the infinitude of the primes. However, whilst in Euler's proof we worked with this one particular sum, in order to prove Dirichlet's theorem, we need to consider a corresponding function for each Dirichlet character. Thus we associate an *L*-function with each Dirichlet character χ as follows:

Definition 3.17. *The L-function*[11] *associated with the Dirichlet character χ is defined as follows:*

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

Each of these *L*-functions converge for $\Re(s) > 1$. And just as $\sum_{n=1}^{\infty} \frac{1}{p^s}$ can be written as a product via the Euler product formula, so each $L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$ can be written as a similar product. The proof, which we omit, relies on the fact that the Dirichlet characters are completely multiplicative, as shown in the previous section.

Theorem 3.18. *Let χ be a Dirichlet character modulo q. Then the L-function associated with χ has an Euler product expansion for $\Re(s) > 1$, i.e.*

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

Now that we have been able to introduce *L*-functions, let us sketch the modern proof of Dirichlet's theorem.

3.2.4 Outline of the proof

Recall that we want to prove that there are infinitely many primes p such that $p \equiv a \pmod{q}$ where a and q are any coprime natural numbers. As in the proof that $\sum_p \frac{1}{p^s}$ 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 q. Indeed, throughout the proof, when we talk about Dirichlet characters, we will be talking about Dirichlet characters modulo q, since they will allow us to isolate the primes number p such that $p \equiv a \pmod{q}$. Thus we obtain:

[11] By analytic continuation, each of these functions except for $L(s,\chi_0)$ can be extended to analytic functions on the domain $\sigma = \Re(s) > 0$

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

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

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

But the terms on the right hand side of the equation for $m > 1$ are negligible; just the order of a constant. So we obtain:

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

The next step allows us to "pick out" the primes in the residue class that we are focused on. We multiply each side of the above equation by $\overline{\chi(a)}$ and then take the sum of these over all the Dirichlet characters modulo q. We should remark that there is another slight "abuse" of notation here, since although we are summing over all of the Dirichlet characters, we will write the sum as being over $U(\widehat{\mathbb{Z}/q}\mathbb{Z})$. There is no harm done by this, since for any n coprime to q, $\chi(n) = \psi(n \pmod q)$ where ψ is the group character that χ is associated with. Thus we have:

$$\sum_{\chi \in U(\widehat{\mathbb{Z}/q}\mathbb{Z})} \overline{\chi(a)} \log L(s,\chi) = \sum_{\chi \in U(\widehat{\mathbb{Z}/q}\mathbb{Z})} \overline{\chi(a)} \sum_{p \nmid q} \frac{\chi(p)}{p^s} \; + \; O(1). \tag{1}$$

Now, we simplify this by first of all interchanging the limits of the series on the right hand side of the above equation, which is permissible because it is absolutely convergent. Then, we appeal to corollary 14, to obtain the following, where $\phi(q)$ is the size of the group $U(\mathbb{Z}/q\mathbb{Z})$:

$$\sum_{\chi \in U(\widehat{\mathbb{Z}/q}\mathbb{Z})} \overline{\chi(a)} \log L(s,\chi) = \phi(q) \sum_{p \equiv a \pmod q} \frac{1}{q^s} \; + \; O(1) \tag{2}$$

We have now obtained an equation that is analogous to the one we obtained when considering the reciprocals of the prime numbers. If we can show that $\lim_{s \to 1+} \sum_{p \equiv a \pmod q} \frac{1}{p^s} = \infty$, then we'll have shown that there must be infinitely many summands, i.e. infinitely many primes p such that $p \equiv a \pmod q$. However, to prove this is considerably more complex than to prove the corresponding result $\lim_{s\to 1+}\sum_p \frac{1}{p^s} = \infty$, since instead of having $\log\sum_{n=1}^{\infty}\frac{1}{n^s}$ on the left hand side, we now have the more complicated $\sum_{\chi\,\in\,U\widehat{(\mathbb{Z}/q\mathbb{Z})}}\overline{\chi(a)}\log L(s,\chi)$ and we must prove that as s tends to 1 from above, this tends to infinity.

In order to prove this, we break it down into two results. The first, easier, result is that $L(s,\chi_0)$ has a simple pole at $s = 1$, but we omit the proof here. The second and much more complicated result is that for all characters except the principal character χ_0, $\lim_{s\to 1} L(s,\chi) \neq 0$. To prove the second result, we divide the characters up into classes as follows:

Definition 3.19. *Classification of Dirichlet characters. The Dirichlet characters modulo q are classified into three classes as follows:*

1. *The first class of characters consists solely of the principal character, which takes the value of 1 for all n that are coprime to q (and 0 otherwise).*
2. *The second class of characters consists of all those characters which take only real values (i.e. 0 or* ± 1*) that are not the principal character.*
3. *The third class of characters consists of those characters which take at least one complex value.*

Different methods of proof are required to demonstrate the result in the case when χ is a real character distinct from the principal character (i.e. belongs to the second class above) and when χ is a complex character (i.e. belongs to the third class) above. Thus to prove the theorem, we need to establish the following results:

Theorem 3.20. *Suppose that χ is a real Dirichlet character modulo q that is distinct from the principal character. Then* $L(s,\chi) \neq 0$ *as s tends to 1 from above.*

Theorem 3.21. *Suppose that χ is a complex Dirichlet character modulo q. Then* $L(s,\chi) \neq 0$ *as s tends to 1 from above .*

Establishing the second theorem is not too difficult, but the case where χ is a real character distinct from the principal character is difficult and requires a substantial amount of work. Dirichlet first proved it by appealing to some heavy duty machinery from the theory of quadratic forms, but the complexity involved spurred some mathematicians to seek out an alternative method. We will omit the proofs of both of these theorems here, but Dirichlet's theorem follows immediately.

With the modern presentation of Dirichlet's proof firmly in our minds, in the next section we shall raise some issues concerning the role the group and Dirichlet characters played that will occupy us for the rest of this thesis. In particular, we shall focus on the *reification* of the characters.

3.3 Analysis of the modern proof: reification

We now distinguish two interconnected but separate issues concerning the development of the function concept: its unification and reification. The first issue concerns the expansion of the permissible domains and co-domains of functions to allow total, single-valued dependences on different collections to be recognized as the same kind of thing as the functions used in analysis. The second concerns taking such dependencies (whether or not explicitly identified as "functions") to be *objects* in some sense. Whilst we will, in places, briefly touch upon the unification of the function concept, we shall primarily be concerned with the second: the reification of the function concept.

Let us first make a brief remark about unification: today we have a unified conception of function and recognize dependencies of many different types as being functions. This is reflected in the modern proof, where we have no qualms about characterizing the group characters, acting on a finite abelian group, or the Dirichlet characters, acting on the natural numbers, explicitly as functions. Indeed, we are comfortable working with a conception of "arbitrary" function, allowing dependencies between any two domains that are single-valued and total with respect to these domains to fall under this concept.

Let us now discuss the delicate issue of reification. We will attempt to tackle it by reference to the use of language and permitted methods within a mathematical theory. That is to say, in discussing mathematical objects, we will take the use of language seriously, at face value. Consequently, if the mathematical theory treats certain candidate entities as objects, we will take this seriously and say that the candidates "really are" objects according to that theory and that the theory is committed to them as such.

However, this leaves us with the question "What does it mean for a mathematical theory to treat a candidate mathematical entity as an object?" We suggest that the following questions can help us answer this:

1. Are the candidate objects given their *own definition* along with appropriate (new) terminology to refer to them?

2. Are the candidate objects studied in their own right? Are they made the subject of theorems which enunciate their properties?

3. Is the treatment of the candidate objects independent of their different representations? That is to say, are we able to work with the candidates without having to continually refer to their particular representations? For example, where appropriate, do we treat the candidates *extensionally*?

4. Is it permissible for the candidate objects to appear in the range of *bound variables*? For example, is it permissible to quantify over them or have them appear in the range of a summation or product index?

5. Is it permissible for there to be dependencies on the candidate objects? For example, is it possible to have a function which acts on the candidates?

We suggest that the above questions can help us to clarify and evaluate whether a theory is treating candidate entities as objects. In particular, we claim that if the answer to *each* of the questions is "yes", then it *suggests* that the mathematical theory and its language treats the candidates as objects and is thus committed to them as such. If the answer to *each* of the questions is "no", then we claim it strongly suggests that the theory is not working with the candidates as objects at all, and so is not committed to them in the same way. And if some of the answers are "yes" whilst others are "no", then it seems the theory is perhaps somewhere in between.

Let us now say a little about each of the questions in turn. If the answer to the first two questions is "yes", it indicates that the candidate objects are given the status of at least grammatical objects: for they are given names and appear as the subject in mathematical propositions. Conversely, if the answer to these two questions is "no", it appears that the candidate objects are not even given the status of *grammatical* objects, and as such it is difficult to see how they could be conceived of as objects in any sense.

With respect to the third question, if we do not continually need to refer to the representations of the candidates in our work, we can view the different representations as all depicting the same object: that which remains after we take away all aspects which are particular to any given representation [12]. Conversely, note that if our treatment of the candidates depends upon which representation we choose, then it becomes difficult to conceive of anything "in common" that is depicted by the various representations. And if the representations do not function as if there is something in common that they serve to represent, it seems unlikely that the candidate entities are taken as objects by the theory.

The fourth questions echoes Quine's claim that "to be is to be the value of a variable" [Qui61, 15]. Thus, if we are sympathetic to his views, then our answer to this question will be "yes" if and only if the mathematical theory is committed to the candidate entities as objects. Even if we are not sympathetic to Quine's view, the fourth question still has relevance. For note that we routinely allow numbers, sets, curves, ideals, and so on to be in the range of bound variables in our mathematical theories. Thus, suppose for example that we have already concluded that

[12] We are thankful to Mic Detlefsen for raising this point.

our theory treats numbers as objects. Moreover, suppose the theory allows both numbers and some other candidate entities to appear within the range of bound variables. Then it allows us to treat the candidates on the same level as other mathematical objects (the numbers) in this sense. Hence it provides us with some motivation for thinking that they are being treated as objects. On the other hand, if the theory does not allow the candidate entities to appear within the range of a bound variable, then it treats them differently to other mathematical objects in a significant way, and provides some motivation for thinking that the theory is not committed to the object-hood of the candidates.

Finally, let us consider the fifth question. First we make a clarifying remark. The question is couched in terms of dependencies rather than functions in order to side-step complications arising from the fact that, traditionally, functions acted on (subsets of) the real or complex numbers. Thus, for example, it is plausible for us to conclude that the mathematical theories of the eighteenth century took natural numbers to be objects, whilst at the same time did not allow functions to act upon them. But, that does not mean that they did not allow general *dependencies* on the natural numbers. Indeed, Euler introduced his famous Totient "function" as follows: "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 ϕA" [Eul44]. Thus Euler did not identify this as a function, i.e. as the same sort of thing that was central to analysis, but it is clear that it depends on the natural numbers.

Now let us offer some explanation as to how the fifth question can help us. Suppose, as in our discussion of question 4, that we have already concluded that our theory takes numbers to be objects and moreover that it allows dependencies on them, for example functions that act upon them. Now suppose that the theory also allows dependencies on the candidate objects. Then the theory treats the candidate entities in the same way that it treats other objects in this regard, and so there is some motivation for concluding that the theory also takes the candidate entities to be objects. However, if the theory does not allow there to be dependencies on the candidate objects, then as it treats them differently to other objects, we have some motivation for concluding that the theory does not treat the candidates as proper objects.

What can the five questions listed above help us conclude about the treatment of characters in the modern proof of Dirichlet's theorem? Well, note that, taking the characters to be the candidate entities, the answer to each of the five questions is "yes". In particular:

1. The 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.

2. The group and Dirichlet characters are studied in their own right prior to the main proof: their general properties are enunciated as theorems and proved. For example, we proved the orthogonality relations for the group characters in Theorem (3.12) and we proved that the Dirichlet characters are completely multiplicative in Theorem (3.16).

3. The Dirichlet characters are treated *extensionally*. For example, in Definition (3.19) we classify them in terms of their values.

4. The group and Dirichlet characters appear in the range of bound variables. For example, in the Euler-product Theorem (3.18), we state the result as "The L-function associated with Dirichlet character χ has an Euler product expansion", where χ is an arbitrary, but fixed, Dirichlet character. Additionally, the characters appear within the range of the indices of summation signs. For example, in stating the orthogonality relations (3.12) and their Corollary (3.13) the index of the sum ranges over the characters.

5. $L(s,\chi)$ can be viewed as a function of the character χ.

We should make a number of remarks concerning the above observations. First, we make the exceedingly plausible assumption that modern analytic number theory treats numbers as objects. Thus, the fact that the modern proof quantifies over and has indices ranging over numbers as well as characters suggests that the characters are treated on a par with other mathematical objects in this sense. Thus, as we argued above, this provides some support for the conclusion that the modern proof treats the characters as objects. The significance of the first three observations is that when we introduce or talk about the characters, we do not have to refer back to a way of constructing them, i.e. we don't have to refer back to products of roots of unity. We did indeed see that the characters can be constructed in this way in Theorem (3.10), but we do not generally keep this in mind when we prove other theorems or utilize the characters in the main proof of Dirichlet's theorem. Specifically, note that the characters are the *subject* of the general theorems: the theorems are not couched in terms of the components of the characters, such as the roots of unity that they can be constructed out of. And indeed, as we will see, it is possible to express the theorems in this way. Similarly, in classifying the characters, we do not opt to classify them according to their composition of products of roots of unity: we instead work with just the values that they take. Thus the modern proof introduces a *new language* to talk about characters even though it could reduce talk about the characters to talk about their components. Thus, the affirmative answers to the five questions provide compelling grounds to conclude that the modern proof treats the characters as objects.

To summarize, we have seen that the modern proof recognizes characters as functions and is also committed to them as *objects*. The earlier presentations of Dirichlet's theorem, however, took quite a different approach. We shall return to issues concerning reification after discussing the earlier proofs, since the contrast between the modern and old approaches will highlight their importance.

4 Dirichlet's Original Proof

4.1 Introduction

In this section, we will sketch Dirichlet's original proof as presented in his 1837 paper [Dir37a]. We will then draw attention to significant features of his approach and subject them to a philosophical analysis. First, however, we will make some very brief remarks about the historical context. Euler was perhaps the first to apply techniques from analysis to solve number theoretic problems, but Dirichlet is often cited as the founder of analytic number theory. He receives this credit in part for his 1837 paper (see e.g. [IK04, 1]) and the methods that spawned from it. Thus the methods Dirichlet employed in this paper were new and particularly ground-breaking at the time, especially considering the importance of his theorem. For, as Dirichlet noted, the result had applications to other theorems (see [DS08, 1]) and, in particular, Legendre's proof of quadratic reciprocity (see e.g. [Rog74]). Let us now consider Dirichlet's proof.

4.2 Sketch of Dirichlet's proof

As we will present the sketch of Dirichlet's proof in a way that remains as faithful to the original as is reasonable, we should make two remarks concerning its organization and some notational conventions. Regarding the organization of the proof, Dirichlet split it into two cases, according to whether the common difference of the arithmetic progression is an odd prime p or not[13]. Dirichlet himself explained the reason for dividing the proof in this manner: "With the novelty of the applied principles it appeared useful to me to start with the treatment of the special case where the difference of the progression is an odd prime, before proving the theorem in its entire generality" [DS08, 2]. Thus in sketching his proof, we will consider both of these cases. Regarding notational conventions, we will follow Dirichlet and use m for the first term of an arithmetic progression, p for the common difference when it is an odd prime, and k for the common difference when it is not an odd prime. This is slightly different from the notation we used in the modern proof, and will again

[13] Dirichlet reduced the latter case to one in which the common difference is divisible by 8.

differ slightly from the later proofs we present. We should also note that when speaking of arithmetic progressions in what follows, we will assume that their first term and common difference are coprime.

4.2.1 Dirichlet's characters and L-functions for an odd prime p

In order to understand Dirichlet's approach to the characters and L-functions we must apply results from modular arithmetic. In particular, we will need to appeal to *primitive roots modulo p* and *indices*. We begin with the definition of these terms and an existence theorem:

Definition 4.1. *Primitive roots modulo p and indices: If p is an odd prime, then any number c which has the property that for any number n which is coprime to p, there is a number γ_n such that $c^{\gamma_n} \equiv n \pmod{p}$ is called a primitive root modulo p. The number γ_n such that $0 \leq \gamma_n < p - 1$ is called the index of n to the base c modulo p.*

Theorem 4.2. *If p is an odd prime, then there exists a primitive root modulo p.*

We will omit the proof, but include an example. Consider the case when $p = 11$. Then 2 will be a primitive root modulo 11. As all natural numbers coprime to 11 are congruent to 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 modulo 11, it suffices to show that for each $m = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ there is an exponent γ_m such that $2^{\gamma_m} \equiv m \pmod{11}$. And indeed the relevant exponents are listed in the table below.

Table 2: 2 is a primitive root modulo 11

Exponent γ_m	0	1	2	3	4	5	6	7	8	9
$m \equiv 2^{\gamma_m} \pmod{11}$	1	2	4	8	5	10	9	7	3	6

The indices modulo p behave similarly to the logarithm[14]. Thus they satisfy the following properties (see e.g. [Wag03, 90]):

Lemma 4.3. *If p is an odd prime and c a primitive root modulo p, then* $\gamma_1 = 0$

Lemma 4.4. *If p is an odd prime and c a primitive root modulo p, then for any m and n coprime to p*

$$\gamma_{mn} \equiv \gamma_m \gamma_n \pmod{p-1}$$

where γ_m is the index of m to the base c modulo p.

[14]In fact, an alternative name for indices is "discrete logarithm" (see e.g. [Wag03, 89]).

Lemma 4.5. *If p is an odd prime, c a primitive root modulo p, then for any m coprime to p and n then*

$$\gamma_{m^n} \equiv n\gamma_m \pmod{p-1}.$$

Dirichlet used primitive roots modulo p and indices in order to construct his L-functions and in what follows we adhere to his presentation closely (see [DS08, 2-4]). Considering an odd prime p, Dirichlet chose a primitive root c of p and a $p-1$th root of unity ω. Then for any q a prime that is distinct from p and s a real variable greater than 1, he considered the geometric series:

$$\frac{1}{1-\omega^{\gamma}\frac{1}{q^s}} = 1 + \omega^{\gamma}\frac{1}{q^s} + \omega^{2\gamma}\frac{1}{q^{2s}} + \omega^{3\gamma}\frac{1}{q^{3s}} + \dots$$

For each such q, we will have a corresponding series and Dirichlet multiplied them all together. Thus on the left hand side, we have

$$\prod \frac{1}{1-\omega^{\gamma}\frac{1}{q^s}}$$

with the product over all primes distinct from p. On the right hand side, the general term for $n = q'^{m'}q''^{m''}\dots$ with n coprime to p and such that q', q'' etc. are all distinct is:

$$\omega^{m'\gamma_{q'}+m''\gamma_{q''}+\dots}\frac{1}{n^s}.$$

However, the lemmas we have stated above (Lemmas (4.3), (4.4), (4.5)) allow us to demonstrate that $m'\gamma_{q'} + m''\gamma_{q''} + \dots \equiv \gamma_n \pmod{p-1}$. Subsequently, as ω is a $p-1$th root of unity, we know that $\omega^{m'\gamma_{q'}+m''\gamma_{q''}+\dots} = \omega^{\gamma_n}$. This is a consequence of the fact that, in modern terminology, the characters are completely multiplicative. Hence we have the Euler-product formula:

$$\prod \frac{1}{1-\omega^{\gamma}\frac{1}{q^s}} = \sum \omega^{\gamma}\frac{1}{n^s} = L. \tag{3}$$

This is the way in which Dirichlet wrote the Euler-product formula, and we should note that it is ambiguous: the γ is really γ_q when it occurs in the left hand side of the equation, whereas when it occurs in the right hand side of the equation, it is γ_n. Dirichlet additionally proved that so long as $s > 1$, L will be absolutely convergent and that, for $\omega \neq 1$, "...will be a function of s that remains continuous and finite for all positive values of s" [DS08, 6].

Dirichlet then noted that by choosing a primitive $p-1$th root of unity Ω, it is possible to express each $p-1$th root of unity as a power of Ω and utilized this to introduce a particular notion for his functions. In particular, he wrote:

The equation just found [Equation (3), above] represents $p-1$ different equations that result if we put for ω its $p-1$ values. It is known that these $p-1$ different values can be written using powers of the same Ω when it is chosen correctly, to wit:

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

According to this notation,we will write the different values L of the series of product as:

$$(4) \quad L_0, L_1, L_2, \ldots, L_{p-2}$$

where it is obvious that L_0 and $L_{\frac{1}{2}(p-1)}$ have a meaning independent of the choice of Ω and that they relate to $\omega = 1$ and $\omega = -1$, respectively [DS08, 3]

First we draw attention to Dirichlet's description of the equation as "representing" different equations: we will return to this later on. Further, notice that the constructions $\Omega^{j\gamma_n}$ where $j \in \{0, 1, \ldots, p-2\}$ are Dirichlet's characters. They correspond, in modern terminology, to the group characters on $U(\mathbb{Z}/p\mathbb{Z})$, so that Dirichlet writes $\Omega^{j\gamma_n}$ while we would write something like $\chi_j(n)$. However we should note that Dirichlet's constructions are neither exactly group characters nor Dirichlet characters in the modern sense. They cannot be group characters because they are defined on *natural numbers coprime to p* and *not* residue classes, and they are not Dirichlet characters because they are not defined on all the natural numbers (though if we stipulate that they take the value of 0 for a natural number that is not coprime to p then they would become Dirichlet characters). Hence L_j for $j \in \{0, 1, \ldots, p-2\}$ are the L-functions corresponding to the characters on $U(\mathbb{Z}/p\mathbb{Z})$.

4.3 Proof of Dirichlet's theorem in the case of an odd prime p

Now that we have seen how Dirichlet introduced his characters and L-functions when dealing with an arithmetic progression whose common difference was an odd prime p, we will consider how he used them to prove his theorem in this special case. However, whilst Dirichlet's presentation of the characters is quite different to the modern approach, the structure of his proof is nonetheless similar. Thus, in sketching his demonstration, we will omit some of the details.

As in the modern proof, Dirichlet demonstrated his theorem by showing that the series $\sum \frac{1}{q^{1+\rho}}$ diverges, with the sum being over all primes of the form $\mu p + m$. Dirichlet began by proving the following results, whose proofs we omit:

Lemma 4.6. *"...for infinitely small ρ, L_0 will become ∞ such that $L_0 - \frac{p-1}{p}\frac{1}{\rho}$ remains finite" [DS08, 5]. That is to say, in more modern terminology, the L-function corresponding to the principal character tends to ∞ as $s = 1 + \rho$ tends to 1.*

Lemma 4.7. *"...the limit approached by $\sum \omega^{\gamma}\frac{1}{n^{1+\rho}}$, with the positive ρ becoming infinitely small, and given that ω does not mean the root 1, will be non-zero" [DS08, 8], i.e. in modern terms, the L-functions that do not correspond to the principal character have a non-zero finite limit as $s = 1 + \rho$ tends to 1.*

Then proof of Lemma (4.7) is the hardest part of the proof, and Dirichlet its demonstration into two parts, whose proofs we shall omit:

Lemma 4.8. *As ρ becomes infinitely small, then the limit approached by $\sum \omega^{\gamma}\frac{1}{n^{1+\rho}}$ for $\omega = -1$ is non zero (closely paraphrased, see [DS08, 8-9] for the proof)*

Lemma 4.9. *As ρ becomes infinitely small, then the limit approached by $\sum \omega^{\gamma}\frac{1}{n^{1+\rho}}$ for ω a complex root of unity is non zero (closely paraphrased, see [DS08, 11-13] for the proof).*

As Dirichlet's characters are constructed out of a $p-1$th root of unity and indices, the character constructed from $\omega = -1$ will be the only character that, in modern terminology, is a real, non-principal character[15]. Thus Lemma (4.8), in modern terminology, is the lemma that $L(s,\chi)$ has a non-zero finite limit for real χ as $s = 1 + \rho$ tends to 1. Similarly, when ω is a complex $p-1$th root of unity, the character constructed from it will take at least one non-real value and thus, in modern terms, is a complex character. Hence Lemma (4.9), paraphrased in modern language, is the lemma that $L(s,\chi)$ has a non-zero finite limit for complex χ as $s = 1 + \rho$ tends to 1.

Assuming that the above results have been proved, let us now consider in more detail how Dirichlet applied them to obtain his result. He first obtained an expression for the logarithm of L:

$$\sum \omega^{\gamma}\frac{1}{q^{1+\rho}} + \frac{1}{2}\sum \omega^{2\gamma}\frac{1}{q^{2+2\rho}} + \frac{1}{3}\sum \omega^{3\gamma}\frac{1}{q^{3+3\rho}} + \ldots = \log L \tag{4}$$

and noted that so long as $\omega \neq 1$ then this would converge to a finite limit; otherwise, it would diverge.

The next step was to "...multiply the equations contained in..." ([DS08, 13]) Equation (4) in a particular way. Before describing his procedure, note Dirichlet's choice of wording, describing one equation as containing others: we will return

[15] Here we apply the terminology for Dirichlet characters to Dirichlet's constructions even though, strictly speaking, they are not Dirichlet characters

to this later on. Dirichlet's procedure was as follows: we take the equation that corresponds to one of the $p-1$th roots of unity, say Ω^j, and multiply each side by $\Omega^{-j\gamma_m}$. Dirichlet then added all of these $p-1$ equations together and obtained the following, with the summations running over q, and γ being the index of q [DS08, 14]:

$$\begin{aligned} &\sum(1+\Omega^{\gamma-\gamma_m}+\Omega^{2(\gamma-\gamma_m)}+\ldots+\Omega^{(p-2)(\gamma-\gamma_m)})\frac{1}{q^{1+\rho}} \\ &+\frac{1}{2}\sum(1+\Omega^{2\gamma-\gamma_m}+\Omega^{2(2\gamma-\gamma_m)}+\ldots+\Omega^{(p-2)(2\gamma-\gamma_m)})\frac{1}{q^{2+2\rho}} \\ &+\frac{1}{3}\sum(1+\Omega^{3\gamma-\gamma_m}+\Omega^{2(3\gamma-\gamma_m)}+\ldots+\Omega^{(p-2)(3\gamma-\gamma_m)})\frac{1}{q^{3+3\rho}} \\ &+\ldots \end{aligned} \tag{5}$$

This rather complicated equation can be simplified by the following lemma:

Lemma 4.10. *If h is a natural number, then:*

$$1+\Omega^{h\gamma-\gamma_m}+\Omega^{2(h\gamma-\gamma_m)}+\ldots+\Omega^{(p-2)(h\gamma-\gamma_m)}=\begin{cases}0 & \text{if } h\gamma-\gamma_m\equiv 0 \pmod{p-1}\\ p-1 & \text{otheriwse}\end{cases}$$

Dirichlet stated this without proving it, but we will present a proof loosely following de la Vallée Poussin (see [dlVP96, 14-15]), whose work on primes and arithmetic progressions we will examine later on.

Proof. Suppose first that $h\gamma-\gamma_m\equiv 0 \pmod{p-1}$. Then as Ω is a $p-1$th root of unity, each term in the sum will be 1. Thus, as there are $p-1$ such terms, we know that the sum must have a value of $p-1$. Suppose now that $h\gamma-\gamma_m\not\equiv 0 \pmod{p-1}$. Then notice that we can make use of the geometric series to write:

$$1+\Omega^{h\gamma-\gamma_m}+\Omega^{2(h\gamma-\gamma_m)}+\ldots+\Omega^{(p-2)(h\gamma-\gamma_m)}=\frac{\Omega^{(p-1)(h\gamma-\gamma_m)}-1}{\Omega^{h\gamma-\gamma_m}-1}.$$

But as Ω is a $p-1$th root of unity, $\Omega^{(p-1)(h\gamma-\gamma_m)}-1=0$ and hence we know the value of the sum must be 0. □

Notice that $p-1$ is the order of the group $U(\mathbb{Z}/n\mathbb{Z})$ and so, if we translate Lemma (4.10) into modern language, it corresponds to a special case of the corollary of the orthogonality relations, for any g, h in a group G:

$$\sum_{\chi\,\in\,\widehat{g}}\chi(g)\overline{\chi(h)}=\begin{cases}|G| & \text{if } g=h\\ 0 & \text{if } g\neq h\end{cases}. \tag{6}$$

Let us now see how this allowed Dirichlet to simplify Expression (5). Notice that each summand is of the form

$$\frac{1}{h}\sum(1+\Omega^{h\gamma-\gamma_m}+\Omega^{2(h\gamma-\gamma_m)}+\ldots+\Omega^{(p-2)(h\gamma-\gamma_m)})\frac{1}{q^{h+h\rho}}, h \in \mathbb{N}.$$

Thus we know that this is going to be 0 unless $h\gamma-\gamma_m \equiv 0 \pmod{p-1}$, i.e. $q^h \equiv m \pmod{p}$. Hence by simplifying Expression (5), we obtain [DS08, 14]:

$$\sum\frac{1}{q^{1+\rho}}+\frac{1}{2}\sum\frac{1}{q^{2+2\rho}}+\frac{1}{3}\sum\frac{1}{q^{3+3\rho}}+\ldots \tag{7}$$
$$=\frac{1}{p-1}(\log L_0+\Omega^{-\gamma_m}\log L_1+\Omega^{-2\gamma_m}\log L_2+\ldots+\Omega^{-(p-1)\gamma_m}\log L_{p-2},$$

with the first sum on the left hand side ranging over all primes $q \equiv m \pmod{p}$, the second ranging over all primes q such that $q^2 \equiv m \pmod{p}$, the third ranging over all primes q such that $q^3 \equiv m \pmod{p}$ and so on.

Thus if we now let ρ tend towards 0 (where, as above, $s = 1+\rho$), $\log L_0$ becomes infinite, since we have already seen that L_0 tends to infinity as ρ tends to 0. However, we have also seen that L_h for $h \neq 0$ has a finite, non-zero limit as ρ tends to 0. Thus $\log L_h$ will be finite. Hence we know that the right hand side of Equation (7) will become infinite as ρ tends to 0. Thus the left hand side must also become infinite. However, we also know that $\frac{1}{2}\sum\frac{1}{q^{2+2\rho}}+\frac{1}{3}\sum\frac{1}{q^{3+3\rho}}+\ldots$ remains finite as ρ tends to 0. Thus we must conclude that $\sum\frac{1}{q^{1+\rho}}$ becomes infinite. But, this sum is over all primes q of the form $q \equiv m \pmod{p}$ and thus Dirichlet's theorem is proved.

4.4 Proof of Dirichlet's theorem in the general case

Now that we have seen the proof of Dirichlet's theorem in the special case where the common difference of the arithmetic progression is an odd prime p, let us consider the more general case. We should note first that Dirichlet does not prove all of the results he needed to fully establish the general case in his 1837 paper. Specifically, he does not include the proof that, in modern terminology, the L functions corresponding to a real, non-principal character has a finite, non-zero limit as $s = 1+\rho$ tends to 1. Of this he wrote, "In the originally presented paper I proved this property using indirect and quite complicated considerations. Later however I convinced myself that the same object can be reached otherwise far shorter" [DS08, 22] and thus omitted the proof, instead referring the reader to his works on quadratic forms. We will not focus on the details of this part of his proof, however.

Thus let us begin by considering how Dirichlet extended his construction of characters in the case where k is not an odd prime. This is a more delicate task than it might first seem, since there are no primitive roots modulo 2^λ where $\lambda \geq 3$. Thus let us consider some more facts from modular arithmetic. First, we state an existence theorem:

Theorem 4.11. *There* are *primitive roots modulo 2, 4 and p^π for p an odd prime and π a natural number*[16]. *That is to say, for $r = 2, 4, p^\pi$ there is some number c_r such that for every n coprime to r there is some natural number j such that $n \equiv c^j \pmod{r}$. When working modulo 2, 1 will be a primitive root and all numbers coprime to 2 will have an index of 1. When working modulo 4, -1 will be a primitive root and all numbers n coprime to 2 will have an index α_n such that $0 \leq \alpha_n \leq 1$. When working modulo p^π, if we take some primitive root c, then the index of a number n coprime to p^π, α_n will be such that $0 \leq \alpha_n \leq (p-1)p^{\pi-1}$.*

We omit the proof of this fact, but see Dirichlet's original paper [DS08, 25] for details.

Second, we note that although there are no primitive roots in the case of 2^λ for $\lambda \geq 3$, we can obtain an analogous result using two numbers, instead of just one primitive root. In particular, it can be shown that:

Theorem 4.12. *If 2^λ where $\lambda \geq 3$, then any number n that is coprime to 2^λ can be represented uniquely in the following form:*

$$n \equiv (-1)^{\alpha_n} 5^{\beta_n} \pmod{2^\lambda},$$

where $\alpha_n \in \{0,1\}$ and $0 \leq \beta_n < 2^{\lambda-2}$. The α_n and β_n are called the indices of n with respect to 2^λ.

Again we omit the proof of this result (for details see [DS08, 15-16]), but we will give an example of the representations in the case for $2^4 = 16$. Since any number coprime to 16 will be congruent to one of $\{1,3,5,7,9,11,13,15\}$ modulo 16, it is sufficient to give the representations for these numbers only. They are given in the table below:

Finally, we should notice the following:

Theorem 4.13. *Primitive roots modulo 2, 4 p^π, and (-1) and 5 when working modulo 2^λ for $\lambda \geq 3$ satisfy the analogue of Theorems (4.3), (4.4), (4.5).*

[16]Dirichlet used π in this context as a natural number and so we will do the same, even though it is not standard practice today.

Table 3: Representations of $1 \leq n \leq 15$ coprime to 16 in the form $(-1)^{\alpha}5^{\beta} \pmod{2^{\lambda}}$

n	$n \equiv (-1)^{\alpha}5^{\beta} \pmod{2^{\lambda}}$
1	$(-1)^0 5^0 \pmod{2^{\lambda}}$
3	$(-1)^1 5^3 \pmod{2^{\lambda}}$
5	$(-1)^0 5^1 \pmod{2^{\lambda}}$
7	$(-1)^1 5^2 \pmod{2^{\lambda}}$
9	$(-1)^0 5^2 \pmod{2^{\lambda}}$
11	$(-1)^1 5^1 \pmod{2^{\lambda}}$
13	$(-1)^0 5^3 \pmod{2^{\lambda}}$
15	$(-1)^1 5^0 \pmod{2^{\lambda}}$

With these results in mind, let us consider how Dirichlet extended his construction of Dirichlet characters and L-functions to the case where the common difference k is not an odd prime. Suppose that $k = 2^{\lambda} p^{\pi} p'^{\pi'} \ldots$ where $\lambda \geq 3$ and p, p', … are distinct odd primes. We omit the case in which $\lambda < 3$ since it is simpler, and in proving the general theorem, Dirichlet omits these cases too.

Given such a k, we choose roots of unity corresponding to the powers of primes in the decomposition as follows: Corresponding to 2^{λ} we choose θ a square root of unity and ϕ a $2^{\lambda-2}$th root of unity, corresponding to p^{π} we choose ω a $(p-1)p^{\pi-1}$th root of unity, corresponding to p'^{π} we choose ω' a $(p'-1)p'^{\pi'-1}$th root of unity and so on. Notice that if the index modulo a prime power in the decomposition can take one of σ values, then we will choose a σth root of unity corresponding to that factor. Then Dirichlet considered the geometric series for $s > 1$, with all indices with respect to q:

$$\frac{1}{1 - \theta^{\alpha}\phi^{\beta}\omega^{\gamma}\omega'^{\gamma'} \ldots \frac{1}{q^s}} = 1 + \theta^{\alpha}\phi^{\beta}\omega^{\gamma}\omega'^{\gamma'} \ldots \frac{1}{q^s} + \theta^{2\alpha}\phi^{2\beta}\omega^{2\gamma}\omega'^{2\gamma'} \ldots \frac{1}{q^{2s}} + \ldots \tag{8}$$

Multiplying each of these series corresponding to every prime q distinct from 2, p, p', … and making use of the properties of the indices mentioned in Lemma (4.13), we are able to obtain a result analogous to complete multiplicity and thus we obtain the Euler-product formula:

$$\prod \frac{1}{1 - \theta^{\alpha}\phi^{\beta}\omega^{\gamma}\omega'^{\gamma'} \ldots \frac{1}{q^s}} = \sum \theta^{\alpha}\phi^{\beta}\omega^{\gamma}\omega'^{\gamma'} \ldots \frac{1}{n^s} = L. \tag{9}$$

Thus Dirichlet's construction $\theta^{\alpha}\phi^{\beta}\omega^{\gamma}\omega'^{\gamma'} \ldots$ are what we recognize today as the characters.

Again, restricting to $s > 1$ ensures that the series is absolutely convergent. And Dirichlet proves that, if at least one of the primitive roots of unity is not equal to 1, then the associated L-series will be a continuous function of s and have a finite limit as ρ tends towards 0, where $s = 1 + \rho$ (see [DS08, 18-19]).

As in the simpler case, Dirichlet made use of the fact that every nth root of unity can be expressed in terms of a power of a primitive nth root of unity in order to introduce a particular notation for the L-series. He wrote [DS08, 17-18]:

> The general equation (10) [our Equation (8)], in which the different roots θ, ϕ, ω, ω', ... can be mutually combined arbitrarily, apparently contains[17] a number K of special equations. To denote the series L corresponding to each of the combinations in a convenient[18] way we can think of the roots of each of these equations (9) [the equations $x^2 - 1 = 0, \phi^{2^{\lambda-2}} - 1 = 0, \omega^{(p-1)p^{\pi-1}} - 1 = 0, \omega^{(p-1)p^{\pi-1}} - 1 = 0, \ldots$] expressed as powers of one of them. Let $\Theta = -1, \Phi, \Omega, \Omega', \ldots$ roots suitable for that purpose, then:
>
> $$\theta = \Theta^{\mathfrak{a}}, \phi = \Phi^{\mathfrak{b}}, \omega = \Omega^{\mathfrak{c}}, \omega' = \Omega'^{\mathfrak{c}'}, \ldots$$
>
> where $\mathfrak{a} < 2$, $\mathfrak{b} < 2^{\lambda-2}$ $\mathfrak{c} < (p-1)p^{\pi-1}$, $\mathfrak{c}' < (p'-1)p'^{\pi'-1}$, ... and, using this notation, denote the series L with:
>
> $$L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots}$$
>
> .

Again, take notice of the use of "contains" in the above paragraph—we will come back to this later.

Thus, the constructions $\Theta^{\mathfrak{a}\alpha}\Phi^{\mathfrak{b}\beta}\Omega^{\mathfrak{c}\gamma}\Omega'^{\mathfrak{c}'\gamma'}\ldots$ are the extension of Dirichlet's characters he introduced in the proof of the special case of the theorem. In particular, they correspond to the Dirichlet characters obtained from characters on the group $U(\mathbb{Z}/k\mathbb{Z})$, although note that as in the special case, they are only defined for natural numbers that are coprime to k and so are not strictly group characters or Dirichlet characters in the modern sense. Hence $L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots}$ are L-functions corresponding to the characters on the group $U(\mathbb{Z}/k\mathbb{Z})$.

[17]The German word is "enthält"

[18]The translator Ralf Stephan uses "comfortable" instead of convenient. The original German word is "bequem".

4.4.1 Sketch of Dirichlet's theorem in the general case

Let us now consider how Dirichlet used his characters and L-functions to prove the more general theorem in which k is assumed to be divisible by 8. Again, due to the similarity in structure to the modern proof and the special case considered above, we will omit some of the details of the proof.

Again, Dirichlet's main goal was to demonstrate that the series $\sum \frac{1}{q^{1+\rho}}$, where the sum ranges over all primes $q \equiv m \pmod{k}$, diverges. And, as above, he needed to establish the following:

Lemma 4.14. *The function* $L_{0,0,0,0,\ldots} = \sum \frac{1}{n^{1+\rho}}$ *becomes infinite as* ρ *tends to 0.*

Lemma 4.15. *The function* $L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots}$*, where not all* $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{c}', \ldots$ *are equal to 0, has a finite non-zero limit as* ρ *tends to 0.*

Presumably to help organize the proofs of these results, Dirichlet categorized his L-functions into three different classes depending on the roots of unity used in their construction. Thus he wrote [DS08, 18]:

> The series denoted with $L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots}$, of which the number equals K, can be divided into three classes according to the different combinations θ, ϕ, ω, ω', … of their roots. The first class contains only one series, namely $L_{0,0,0,0,\ldots}$, that is, the one where:
>
> $$\theta = 1,\ \phi = 1,\ \omega = 1,\ \omega' = 1,\ \ldots$$
>
> holds. The second class shall cover all other series with only real solutions to the equations (9)such that therefore to express those series we have to combine the signs in:
>
> $$\theta = \pm 1,\ \phi = \pm 1,\ \omega = \pm 1,\ \omega' = \pm 1,\ \ldots$$
>
> in every possible way excepting only the combination corresponding to the first class. The third class finally includes all series L where at least one of the roots ϕ, ω, ω', … is imaginary, and it is evident that the series of this class come in pairs since the two root combinations:
>
> $$\theta, \phi, \omega, \omega', \ldots; \quad \frac{1}{\theta}\frac{1}{\phi}\frac{1}{\omega}\frac{1}{\omega'}, \ldots$$
>
> are mutually different given the above mentioned condition.

In particular, Dirichlet's first class contains only the L-function corresponding to the principal character, the second class consists of those L-functions corresponding to a real character that is not identical to the principal character, and the third class contains all of the L-functions associated with complex characters.

As above, Dirichlet split his demonstration of Lemma (4.15) into two cases:

Lemma 4.16. *The limit for an L-function L belonging to the second class as ρ tends to 0 where $s = 1 + \rho$ is non-zero.*

Lemma 4.17. *The limit for an L-function L belonging to the third class as ρ tends to 0 where $s = 1 + \rho$ is non-zero (see [DS08, 19-21])*

Assuming that these results have been established, we can proceed in an analogous way to in Dirichlet's proof of the special case of his theorem. In particular, we take logarithms of each side of Equation (9) to obtain for each combination of roots of unity:

$$\sum \Theta^{\mathfrak{a}\alpha}\Phi^{\mathfrak{b}\beta}\Omega^{\mathfrak{c}\gamma}\Omega'^{\mathfrak{c}'\gamma'}\dots\frac{1}{q^{1+\rho}} + \frac{1}{2}\sum \Theta^{\mathfrak{a}2\alpha}\Phi^{\mathfrak{b}2\beta}\Omega^{\mathfrak{c}2\gamma}\Omega'^{\mathfrak{c}'2\gamma'}\dots\frac{1}{q^{2+2\rho}} + \dots$$
$$= \log L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}'}.$$

The next step is to multiply both sides of the above equation by

$$\Theta^{-\alpha_m\mathfrak{a}}\Phi^{-\beta_m\mathfrak{b}}\Omega^{-\gamma_m\mathfrak{c}}\Omega'^{-\gamma'_m\mathfrak{c}'}\dots$$

For comparison, in modern terminology $\Theta^{-\alpha_m\mathfrak{a}}\Phi^{-\beta_m\mathfrak{b}}\Omega^{-\gamma_m\mathfrak{c}}\Omega'^{-\gamma'_m\mathfrak{c}'}\dots$ could be written as something like $\chi_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\dots}(m)$, although we probably would not want to include all the construction data in the subscript.

Next, we take the sum of all these equations, with the sum running over all of the root combinations, from $\mathfrak{a} = \mathfrak{b} = \mathfrak{c} = \mathfrak{c}' = \dots = 0$ to $\mathfrak{a} = 1, \mathfrak{b} = 2^{\lambda-2} - 1, \mathfrak{c} = (p-1)p^{\pi-1} - 1, \mathfrak{c}' = (p'-1)p'^{\pi'-1} - 1, \dots$...

Then, to quote Dirichlet exactly:

> ... the general term on the left hand side becomes:
>
> $$\frac{1}{h}W\frac{1}{q^{h+h\rho}},$$
>
> where the sum is over all primes q and W means the product of the sums taken over $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{c}', \dots$ or respectively over:
>
> $$\sum \Theta^{(h\alpha-\alpha_m)\mathfrak{a}},\ \sum \Phi^{(h\beta-\beta_m)\mathfrak{b}},\ \sum \Omega^{(h\gamma-\gamma_m)\mathfrak{c}},\ \sum \Omega'^{(h\gamma'-\gamma'_m)\mathfrak{c}'},\dots$$
>
> ([DS08, 20])

Fortunately this can be simplified by utilizing, in modern terminology, a corollary of the orthogonality relations for the characters, as we did in the proof of the special case (see Lemma (4.10)). In this case, it is more complex as Dirichlet applied an orthogonality relation for each factor in the decomposition of p. His application was as follows:

> We can now see ... that the first of these sums is 2 or 0, corresponding to if the congruence $h\alpha - \alpha_m \equiv 0 \pmod 2$ or, equally, the congruence $q^h \equiv m \pmod 4$ holds or not; that the second is $2^{\lambda-2}$ or 0 corresponding to if the congruence $h\beta - \beta_m \equiv 0 \pmod{2^{\lambda-2}}$ holds, or equally 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 if the congruence $h\gamma - \gamma_m \equiv 0 \pmod{(p-1)p^{\pi-1}}$ or, equally, the congruence $q^h \equiv m \pmod{p^\pi}$ holds or not, and so on; that therefore W always disappears except when the congruence $q^h \equiv m$ holds modulo each of the modules $2^\lambda, p^\pi, p'^{\pi'}, \ldots$, that is, when $q^k \equiv m \pmod k$ holds, in which case $W = K$ [where K is the order of the group $U(\mathbb{Z}/n\mathbb{Z})$]. Our equation thus becomes:
>
> $$\sum \frac{1}{q^{1+\rho}} + \frac{1}{2}\sum \frac{1}{q^{2+2\rho}} + \frac{1}{3}\sum \frac{1}{q^{3+3\rho}} + \ldots \qquad (10)$$
> $$\frac{1}{K}\sum \Theta^{-\alpha_m \mathfrak{a}}\, \Phi^{-\beta_m \mathfrak{b}} \Omega^{-\gamma_m \mathfrak{c}} \Omega^{-\gamma_{m'} \mathfrak{c}'} \ldots \log L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots},$$
>
> where the summation on the left is over all primes q the first, second, third powers of which are contained in the form $\mu k + m$, while the summation on the right is over $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{c}', \ldots$ and extends between the given limits. [DS08, 20]

And now, the proof of the theorem is nearly complete. For as before, we know that the term in the right hand side of Equation (10) $\log L_{0,0,0,0,\ldots}$ tends to infinity as ρ tends to 0. Thus the left hand side must also tend towards infinity, but given that the terms $\frac{1}{2}\sum \frac{1}{q^{2+2\rho}} + \frac{1}{3}\sum \frac{1}{q^{3+3\rho}} + \ldots$ will be finite, we must have that $\sum \frac{1}{q^{1+\rho}}$ diverges. Thus, the theorem is established.

4.5 Observations and analysis

In this section, we will draw attention to a number of significant aspects of Dirichlet's proof, and contrast it to the modern presentation, before examining the philosophical consequences of these features. In particular, we will focus on the following:

1. The introduction of Dirichlet's characters in both the special and general case.
2. The notation for the L-functions.
3. Dirichlet's techniques and their consequences.

4.5.1 Introduction of Characters

Thus let us begin by considering Dirichlet's introduction of his characters. Recall that, in the modern proof, we define both the group and Dirichlet characters and study them in their own right. The definition focuses on the *general properties* of the characters, and we prove further general theorems about them. For example, we prove the orthogonality relations for the group characters, that the characters on a group G with operation $\cdot$ themselves form a group under multiplication that is isomorphic to the original group G, and that the Dirichlet characters are completely multiplicative. Moreover, we are quite content to refer to an arbitrary character by using the single symbol 'χ'. However, Dirichlet did *not* give his characters their own definition. Rather, they appeared as part of his construction of the L-functions and were not even given a name: nowhere in the paper did Dirichlet refer to his constructions as "characters" or utilize any other term for the purpose. Moreover, whilst the definitions in the modern proof focus on the *properties* that the characters have, Dirichlet's characters were *explicitly constructed*. Furthermore, he continually referred to the characters as Ω^{γ} in the special case and $\Theta^{\alpha}\Phi^{\beta}\Omega^{\gamma}\Omega'^{\gamma'}\dots$ in the general case: he did not allow himself to utilize a single symbol to abbreviate the characters, as we do today. Thus, Dirichlet included a much greater amount of explicit information about the make-up of the individual characters than is given or needed in the modern proof.

Moreover, Dirichlet's treatment of the characters did not include explicitly isolating as theorems the general properties and relations that they satisfy as *properties which they bear*. For example, whereas we express complete multiplicity in Lemma (3.16) in terms of two properties which all Dirichlet characters satisfy, in the special case Dirichlet simply wrote, "But now it holds that $m'\gamma_{q'} + m''\gamma_{q''} + \dots \equiv \gamma_n \pmod{p-1}$ and because [Ω is a $p-1$th root of unity] $\omega^{m'\gamma_{q'}+m''\gamma_{q''}+\dots} = \omega^{\gamma_n}$" [DS08, 2] (recall the discussion around Equation (3)). This is thus presented by Dirichlet as a result or a calculation rather than a property that can be explicitly attributed to his constructed characters. Moreover, in the general case, Dirichlet did not even state the result in this way. Rather, he referred to "...the above mentioned properties of indices and equations" [DS08, 17]. Thus he referred back to properties of the *components* of the characters, but would *not* refer to properties of the characters themselves. Similar features can be seen in Dirichlet's treatment of the orthogonality relations in the general case. For he did not state them as relations or properties which hold of the characters, but instead performed a calculation to evaluate the sum of each component in the product W (recall the discussion of [DS08, 20], above). However, he did not isolate the general relation that unifies all of these calculations, which we expressed in modern notation as Corollary (6).

4.5.2 Notation for L-functions

The above differences between the modern and original proof are also reflected in the notation for the L-functions. For, in the modern proof, recall that we associate to each χ an L-function $L(s,\chi)$, i.e. the L-function is specified by indicating a *function* that it corresponds to. And, moreover, we can think of $L(s,\chi)$ as a *function of a function* χ. In Dirichlet's proof, however, we saw that the L-functions were specified by *numbers* that *described the construction of the characters*. Thus, in the simpler case, fixing a primitive root c of the difference p and a primitive $p-1$th root of unity Ω, the exponent h in Ω^h provided a means of identifying which root was used in the construction of the characters. Thus, Dirichlet could identify the corresponding L-series by reference to h, suppressing the choice of c and Ω, as L_h. Similarly, in the general case, Dirichlet fixed primitive roots corresponding to the odd prime powers in the factorization of the common difference, including (-1) and 5 when it was divisible by 8, and corresponding primitive roots of unity associated with these factors. This, as we saw, allowed him to specify which roots were used in the construction of his characters as $\Theta^{\mathfrak{a}}$, $\Phi^{\mathfrak{b}}$, $\Omega^{\mathfrak{c}}$, $\Omega'^{\mathfrak{c}'}$, ... This allowed for the corresponding L-function to be identified in terms of the exponents $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$, $\mathfrak{c}'$, ..., again suppressing the choice of the primitive roots and primitive roots of unity, as $L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots}$. Thus it appears that Dirichlet did not share the modern conception of $L(s,\chi)$ as a function of characters.

We should also remark upon Dirichlet's classification of the L-functions. As we saw in the quote in Section (4.4.1) ([DS08, 18]), Dirichlet's division was based upon the values of the *roots of unity* that appear in the construction of the L-functions. Thus, as this classification is based upon the description of how the characters are composed, it relies on an *intensional* classification of the characters. The modern proof, on the other hand, uses an *extensional* classification of the characters, based only on the *values of the characters*. Thus, again, Dirichlet's approach to the L-functions provided much greater information about their make-up than is necessary in the modern proof.

4.5.3 Dirichlet's methods and techniques

We have highlighted above that Dirichlet's introduction of the characters and L-functions provided much more information about the make-up of the characters and L-functions than the modern proof does. And, understandably, his proof also utilized different techniques and operations in dealing with these constructions. In particular, there are certain operations that we are willing to undertake in the modern proof which Dirichlet did not use.

The first such operation is quantification. Recall that in the proof of the simple case, after obtaining the Euler-product formula, Dirichlet remarked that it "...represents $p-1$ equations that result if we put for ω its $p-1$ values" [DS08, 3] and that he later instructed the reader to "... multiply the equations contained in ..." Equation (4) (the series expansion of the logarithm of L). And, in the proof of the more general result, Dirichlet noted that Equation (8) "...apparently contains a number K of special equations" [DS08, 17]. This is subtly different from the approach taken in the modern proof. For recall that in the statement of the Euler-product formula, and when considering the the expansion of the logarithm of L, the modern proof involves *quantification* over the characters. Consequently, in the modern proof, the places in which Dirichlet made references to containment or representation are replaced by quantification. It subsequently appears that Dirichlet was reluctant to quantify over his constructed characters in the same way that we do today.

Related to Dirichlet's unwillingness to quantify over characters was his avoidance of sums or products which have *indices ranging over characters*. For note that in the modern proof, we do make use of such sums: the statement of the orthogonality relations (see Equations (3.12), (3.13)) and the equations obtained via multiplying the series expansion of $\log L(s,\chi)$ and then summing the result (see Equations (1), (2)) all have indices explicitly ranging over the characters. Dirichlet, however, adopted a different approach. As we have seen in Equation (4.10), his application of the orthogonality relations in the special case of the proof was *not* written using a summation sign, but by writing out the whole sum, modulo the ellipsis. He adopted the same approach when multiplying and summing the series expansions of the logarithm of L in the proof of the special case (see Equations (5) and (7)). Thus, Dirichlet avoided having to specify an index for the summation. And, in the proof of the more general theorem, he struggled to express his application of the orthogonality relations. Indeed, as we have seen, he expressed it mostly in words, and had to do so *for each of the factors* in the prime decomposition of the common difference. Where he did attempt to express it in an equation, he was forced to use the single symbol W which stood for a product of sums taken over the *numbers* used to specify the characters, *not* the characters themselves. And when having to address the more complicated multiplication and sum of the series expansion of $\log L$ in the more general case, Dirichlet made use of his characterization of the L-functions in terms of these numbers once more, using a summation sign that once again ranged over the numbers which appeared as exponents in the construction of the characters (see Equation (10)).

The unwillingness to have sums whose indices ranged over the characters also provides further evidence that Dirichlet could not view the L-function as a function of his constructed characters in their entirety. For notice that, in the modern proof, when we write an expression such as

$$\sum_{\chi \in \widehat{U(\mathbb{Z}/q\mathbb{Z})}} \overline{\chi(a)} \log L(s,\chi) = \phi(q) \sum_{p \equiv a \pmod{q}} \frac{1}{q^s} + O(1)$$

we really are emphasizing the fact that $L(s,\chi)$ depends on χ because it is functioning as the index. That is to say, it emphasizes that this χ is a variable that is to be replaced by each of the characters in turn. But, when Dirichlet instead wrote:

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

for the special case and:

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

in the general case, this emphasis is not present. In neither of these equations do we have any emphasis on the entire construction $\Omega^{h\gamma}$ or $\Theta^{\alpha\mathfrak{a}}\, \Phi^{\beta\mathfrak{b}} \Omega^{\gamma\mathfrak{c}} \Omega^{\gamma'\mathfrak{c}'}\ldots$ being replaced within the L-function, because, for Dirichlet, *it does not vary*. Rather, we have already assumed that Ω is fixed, where γ is the index, and so all that *can* vary is the exponent: a number.

4.5.4 Analysis

In the above, we have highlighted various features of Dirichlet's proof and contrasted them with the modern approach. We will now consider the consequences of these aspects of Dirichlet's presentation and what they tell us about his conception of character and function.

Let us first examine what we can conclude about Dirichlet's treatment of the characters. We first of all make a brief note about the issue of the unification of the function concept. At this point in history, it may be that mathematicians did not view total, single-valued dependencies on the natural numbers as the same sort of thing as the functions of analysis. Indeed, we found almost no references to such dependencies as functions prior to 1850, when Eisenstein explained how it was that these dependencies could be identified as functions. He wrote: "For the concept of function, one moved away from the necessity of an analytic connection, and began to view its essence (of that concept) in the tabular "composition" of a row of values associated with the values of one or several variables. Thus, it became possible, to categorize those functions under the concept that-due to "conditions" of an arithmetic nature-receive 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 either indeterminate and arbitrary or without any meaning."[19] [dWzB50]. And, given that Dirichlet did not give the characters themselves a name or explicitly identify them as "functions", it may be that he did not view them as the same type of dependencies as those studied in analysis.

Let us now consider reification. We draw attention to the following points:

1. Dirichlet introduced the characters only as an auxiliary construction in his definition of the L-functions and referred to them throughout the proof by describing their construction.

2. Dirichlet did not present a study of the general properties of the characters independently of their appearance in his main proof. Indeed, it appears that he did not want to attribute certain general properties to the characters at all. Instead, the theorems were couched in terms of the properties of the *components* of the characters.

3. Dirichlet's characterization of L-functions relied on an *intensional* classification of the characters, i.e. on its representation as a product of roots of unity. Dirichlet did not permit quantification over characters-he quantified over part of their construction data instead.

[19] Translation by Wilfried Sieg. The original German is: 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 oder der (mehrerer) Variabeln zu setzen anfing, war es möglich, auch solche Funktionen unter diesen Begriff mit aufzunehmen, welche aus Bedingungen arithmetischen 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.

4. Dirichlet did not permit summation indices to range over the characters. Again, the indices had to range over part of the construction data, which were numbers.

5. Dirichlet did not conceive of his *L*-functions as a function of his constructions $\Theta^{\alpha\mathfrak{a}}\ \Phi^{\beta\mathfrak{b}}\Omega^{\gamma\mathfrak{c}}\Omega^{\gamma\mathfrak{c}'}\ldots$

Let us elaborate upon the observations listed above. First we again assume that Dirichlet treated the natural numbers as mathematical objects, as seems eminently plausible. And note that he did indeed allow summation indices to range over numbers, and allowed for there to be *dependencies* on the natural numbers. Indeed, he was willing to have summation indices which ranged over numbers in the construction data for his characters, and his *L*-functions were described in terms of this data, and as such a dependency on the natural numbers. Thus, for reasons we discussed in Section (3.3), the fourth point suggests that Dirichlet did *not* treat the characters as objects. The first three points also strongly indicate that Dirichlet viewed the characters not as something independent in their own right, but as nothing more than the product of roots of unity and associated indices. Specifically, the fact that the construction of the characters appeared only as an auxiliary in a different construction and were neither defined nor studied independently gives the impression that the components of the characters do not form anything together independently of the larger construction of the *L*-functions. This is reiterated both by Dirichlet's apparent reticence to attribute general properties to the characters when he was willing to attribute properties to their components and his intensional characterization of the *L*-functions in terms of the characters. For these features suggests that the components of the characters (i.e. the roots of unity, the indices, and the numbers used to distinguish between them) are legitimate mathematical objects, but their composition as characters is not something independent, it is just a part of the construction of the *L*-functions. The intensional classification further implies that the characters are completely *dependent* upon their description in terms of their components. Thus, we see that Dirichlet operated in exactly the opposite way to the modern proof: he reduced all talk of the characters to talk of their components and in doing so provided much more explicit information about their construction. Thus all of the answers to the five questions listed in Section (3.3) are "no" and as such we have strong reason to assert that Dirichlet did not treat his characters as objects.

Let us now consider the consequences that Dirichlet's approach had on the proof. First note that a number of features of Dirichlet's approach means that his organization of the proof is quite distinct to the modern one. In particular, the fact that Dirichlet did not give the characters their own independent definition and introduced them only as an auxiliary *construction*, and the fact that he did not present a separate study of their general properties, makes his proof more difficult for us to follow. For although the construction provides us with details concerning the nature of each individual character, it does not immediately yield crucial general properties, such as the result analogous to complete multiplicity or the applications obtained via the orthogonality relations. And, as the general properties of the characters are not presented in an independent study prior to the proof, or even enunciated as general properties within the proof itself, when we reach a step in the argument that relies upon them, we face an added difficulty. For we must now pause and consider how to obtain the application from the constructed definition, before we can return our attention back to the main proof. Thus, as applications of general results about characters are obtained in the middle of the main proof without a statement of the general result they follow from, our attention is switched back and forth between following derivations of these applications and the main proof itself. This could be reduced if, as in the modern approach, the characters had been presented and their general properties stated and derived before the main proof began, so that they could be referred to at the necessary points.

Secondly, we note that, combined with the lack of summation signs whose indices range over the characters themselves and a conception of the L-functions that did not present them as depending on the characters, the additional information that we have about the characters can appear to be somewhat overwhelming and repetitive to the modern reader. For in order to establish applications of e.g. complete multiplicity and the orthogonality relations, the proof had to account for each of the factors in the prime decomposition of the common difference of the arithmetic progression. Consequently, Dirichlet stated (and proved) numerous special cases of the applications that he wanted to obtain, as we highlighted above.

Moreover, the construction information became unwieldy and difficult to handle in the proof. Indeed, as we noted above, it is difficult to express the application of the orthogonality relations in the general proof. Recall that Dirichlet considered only the general term of the left hand side of the equation, which he wrote as $\frac{1}{h}\sum W\frac{1}{q^{h+h\rho}}$ with W denoting the product of the various sums

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

The symbol 'W' provides us with no information about the structure of the product it is supposed to denote. And the right hand side of the equation looks unwieldy: $\frac{1}{K}\sum \Theta^{-\alpha_m\mathfrak{a}}\ \Phi^{-\beta_m\mathfrak{b}}\Omega^{-\gamma_m\mathfrak{c}}\Omega^{-\gamma_{m'}\mathfrak{c}'}\ldots\log L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}'\ldots,}$. In the modern proof, we allow ourselves to have sums whose indices range over characters and we treat $L(s,\chi)$ as a function of χ. These two features compliment each other nicely. By thinking of $L(s,\chi)$ as depending on χ, we can consider what happens as χ varies over all characters in the group. Then by allowing a summation whose index ranges over characters, we can express succinctly and perspicuously what Dirichlet struggled with in Equation (1):

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

This highlights the relationship between the L-function $L(s,\chi)$ and the character its logarithm is multiplied by: the L-function is multiplied by the conjugate of the character it corresponds to. Unfortunately, Dirichlet's notation tends to obscure this relationship: for the logarithm of the appropriate L-function $L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}',\ldots}$ is multiplied by the character $\Theta^{-\alpha_m\mathfrak{a}}\ \Phi^{-\beta_m\mathfrak{b}}\Omega^{-\gamma_m\mathfrak{c}}\Omega^{-\gamma_{m'}\mathfrak{c}'}\ldots$. Thus we have to sift through the roots of unity to notice the exponents $-\mathfrak{a}, -\mathfrak{b}, -\mathfrak{c}, -\mathfrak{c}', \ldots$ and compare them to the subscripts of the L-function $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{c}', \ldots$.

Consequently, we see that the presentation of characters and L-functions in Dirichlet's original proof arguably impedes our ability to follow and understand the proof in at least three separate ways:

1. The general properties of the characters are not stated.

2. Numerous applications of special cases of results need to be obtained in order to obtain the more general application of the result.

3. The notation is overwhelming and unable to marshal the information that Dirichlet included.

4.5.5 Summary

We have thus seen that Dirichlet did not treat the characters as independent mathematical objects and may not have viewed them as functions, in the same category as functions from analysis. We have also seen that the presentation of characters and L-functions in Dirichlet's proof can impede the understanding of a modern reader. However, as we come to examine the works of later mathematicians, we will see that they gradually moved away from the features of Dirichlet's approach listed above, as well as towards taking functions of characters and recognizing characters as functions. Thus they moved towards treating characters as both functions and objects.

5 Transitional Presentations: Dedekind, de la Vallée Poussin, Hadamard and Landau

As we have now seen both a modern presentation of Dirichlet's proof and the original, we will not provide an outline of the proofs given by Dedekind, de la Vallée

Poussion, Hadamard[20] and Landau. Instead, we will focus on how the various presentations changed with respect to the treatment of the characters and Dirichlet functions.

5.1 Dedekind's presentation

5.1.1 Description of Dedekind's use of characters and L-functions

We start with Dedekind's presentation found in Supplement *VI* of the *Vorlesungen über Zahlentheorie*. In this supplement, Dedekind established the theorem on primes in arithmetic progressions in its full generality. He used m for the first term of this arithmetic progression, k for the common difference which was assumed to be coprime to m and generally used q for a prime number. His presentation followed Dirichlet's original proof quite closely, although there are some important differences between the two.

First, Dedekind presented an overview of the important steps needed to establish the theorem in §123: in particular, he argued that the Euler-Product formula holds and obtained the series expansions for the L-functions and their logarithms. However, this overview was presented for a more general class of L-functions than we have considered previously. Indeed, Dedekind introduced an extended class of L-functions as follows:

> 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 $\psi(1) = 1$ [DD99, 237]

Note that the condition that $\psi(n)\psi(n') = \psi(nn'),\ \psi(1) = 1$ amounts to complete multiplicity.

Dedekind specialized his general class of L-functions to Dirichlet functions in the very next section. In order to introduce them, he first constructed the characters

[20] In his paper *Sur la distrbution des zéros de la fonction $\zeta(s)$ et ses conséces arithmétiques*, Hadamard does not re-prove Dirichlet's theorem in its entirety, but considers key steps involved in the proof whilst examining other, related theorems.

as Dirichlet had[21], and still refrained from calling them "characters". However, there were some significant changes to his presentation. In particular, Dedekind introduced the characters separately, before introducing the L-series in which they feature. He was also willing to introduce functional notation to represent them, setting $\phi(n) = \frac{\theta^\alpha \eta^\beta \omega^\gamma \omega'^{\gamma'} \dots}{n^s}$, and $\chi(n) = \theta^\alpha \eta^\beta \omega^\gamma \omega'^{\gamma'} \dots$. Further Dedekind noted that the characters were completely multiplicative and attributed this to them as a property: "The numerator [of $\phi(n)$] $\chi(n) = \theta^\alpha \eta^\beta \omega^\gamma \omega'^{\gamma'} \dots$ has the *characteristic property* $\chi(n)\chi(n') = \chi(nn') \dots$" [DD99, 239 emphasis added]. Finally, whereas Dirichlet's characters had only been defined for natural numbers coprime to the common difference, Dedekind implied that his characters could be construed as ranging over all the natural numbers. For note that the ψ occurring in his general L-function was defined for *all* natural numbers and at the end of the introductory section, Dedekind remarked"... if the value of the function ψ is 0 for all primes dividing a particular number k, then $\psi(n) = 0$ for any n not relatively prime to k, and the equations [the Euler-Product formula and the series expansion for the logarithm of L] remain correct when one allows n to run through all numbers relatively prime to k and q through all primes not dividing k" [DD99, 238]

With the characters suitably defined, Dedekind then introduced the Dirichlet L-functions:

> ... we can assert the equations (I) and (II) of the previous section:
>
> $$\prod \frac{1}{1 - \psi(q)} = \sum \psi(n) = L$$
>
> $$\sum \psi(q) + \frac{1}{2} \sum \psi(q^2) + \frac{1}{3} \sum \psi(q^3) + \dots = \log L$$
>
> in which q runs through all primes not dividing k and n runs through all numbers relatively prime to k. As long as $s > 1$, both series have sums independent of the order of terms. [DD99, 239].

And, echoing similar turns of phrase from Dirichlet, Dedekind claimed "... 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

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

different particular series" [DD99, 240, emphasis added]. He then proceeded to divide the characters into classes. Like Dirichlet, Dedekind described each of the

[21] While Dirichlet expressed each root of unity composing the character as a power of a primitive root of unity, Dedekind did not.

three classes that a Dirichlet L-function can belong to in terms of the roots of unity that feature in the construction of its corresponding character. He wrote:

> We divide these series L into three classes:
>
> In the *first* class there is only one series L_1, namely the one for which all the roots of unity θ, ν, ω, ω',... have the value 1.
>
> In the *second* class we include all the remaining series L_2 for which the roots of unity are real, and hence equal to ± 1.
>
> In the *third* class we include all remaining series L_3, that is, those for which at least one of the roots of unity is imaginary. The number of these series is even, since they can be grouped in conjugate pairs—if one such series L_3 corresponds to the roots θ, ν, ω, ω',..., then there is a second series [which Dedekind called L_3'] corresponding to the roots θ^{-1}, ν^{-1}, ω^{-1}, ω'^{-1},..., and these two systems of roots are not identical [DD99, 240].

Dedekind utilized this classification and associated notation when dealing with summations of L-functions in a way that is different to the modern presentation. For example, when proving that the L-functions corresponding to a complex character have a finite non-zero value as s tends to infinity, he obtained the following equation:

$$\phi(k)(\sum \frac{1}{q} + \frac{1}{2}\sum \frac{1}{q^{2s}} + \ldots + \frac{1}{\mu}\sum \frac{1}{q^{\mu s}} + \ldots)$$
$$= \log L_1 + \sum \log(L_2) + \sum \log(L_3 L_3'),$$

> 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" [DD99, 245]

Thus Dedekind here opted to allow the L-functions themselves to appear as the value of a bound variable, instead of the characters. In contrast, the modern presentation writes the same equation with the summation indices ranging over the characters themselves, as following (altering slightly the symbols for the common difference of the arithmetic progression, the variable, and primes to make it consistent with Dedekind's use): [22]

$$\sum_{\chi\in\widehat{G}} \log L(s,\chi) = \sum_{\chi\in\widehat{G}} \sum_{q\nmid k} \sum_{n=1}^{\infty} \frac{\chi(q)^n}{np^{sn}}.$$

Another approach that Dedekind took to the sums of L-functions was to have the summation indices ranging over the components of the characters, in a similar way to Dirichlet. However, in some cases he was able to shorten his expressions by means of notational abbreviations that Dirichlet did not make. For example, given a particular collection of roots of unity $\theta, \eta, \omega, \omega', \ldots$, Dedekind denoted $\theta^{-\alpha_1}\eta^{-\beta_1}\omega^{-\gamma_1}\omega'^{-\gamma'_1}\ldots$ by χ, where $\alpha_1, \beta_1, \gamma_1, \gamma'_1$ denoted the indices of the first term of the progression m. Note that whilst Dedekind used the symbol $\chi(n)$ as a function symbol for the characters in a footnote, he explicitly called χ, as defined here, a *value*, and indeed it is $\chi(m)$. Moreover, when he utilized it in a summation, Dedekind was explicit that the summation ranges not over χ, but rather the roots of unity it contains:

> The summation of all products $\chi \log L$ therefore gives the result
>
> $$\phi(k)(\sum \frac{1}{q^s} + \frac{1}{2}\sum \frac{1}{q^{2s}} + \frac{1}{3}\sum \frac{1}{q^{3s}} + \ldots) = \sum \chi \log L,$$
>
> 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 $\phi(k)$ different root systems $\theta, \eta, \omega, \omega', \ldots$ [DD99, 247]

Dedekind's approach was not quite the same as Dirichlet's, since Dedekind's summation indices ranged over systems of roots of unity, whereas Dirichlet's ranged over exponents that the primitive roots of unity were raised to. This is more abstract than Dirichlet's approach, as Dirichlet had to choose a representation of the characters in order to sum over their indices, but Dedekind did not have to do this. However, neither Dirichlet nor Dedekind allowed the summation indices to range over the characters themselves, but rather only over certain *numbers* used in their construction. Note moreover that Dedekind's procedure, like Dirichlet's, somewhat obscures the operations that are involved in the expression. For in the modern proof, the relationship between $\widehat{\chi(m)}$ and the χ occurring in the argument position of $L(s,\chi)$ is made explicit and is easily visible. Within Dedekind's expression,

[22] The equation appears in the proof of Theorem (3.21), although we did not include the proof in chapter 3. See e.g. [EW05, 220-221]

however, this relationship is not made explicit. Indeed, there is nothing in the notation χ to directly reflect the fact that it is dependant upon the character and the first term of the arithmetic progression.

Finally, we should note that whilst Dedekind was willing to attribute the property of complete multiplicity to characters, he did not formulate the orthogonality relations or Corollary (3.13) in the same way. Indeed, these were formulated so that the characters were not the subject of the theorems at all: they were instead formulated in terms of the roots of unity that featured in their construction. The two orthogonality relations and their corollary are given below[23]:

Theorem 5.1. *(Half of the First Orthogonality Relation) Let* $f(x) = \sum \theta^{\alpha}\eta^{\beta}\omega^{\gamma}\omega'^{\gamma'} \dots x^{\nu}$ *with the summation being over the integers* ν *that are less than and coprime to* k*. Then* $f(1) = 0$ *(see [DD99, 241]).*

Theorem 5.2. *(Second Orthogonality Relation) Let* q *be a prime that is not divisible by* k*, and* μ *a positive integer. Then the product*

$$\sum \theta^{\alpha\mu} \sum \eta^{\beta\mu} \sum \omega^{\gamma\mu} \sum \omega'^{\gamma'\mu} \dots,$$

where the summation ranges over the respective $a, b, c, c', \dots$ *values of* $\theta, \eta, \omega, \omega', \dots$*, is non-zero with value* $\phi(k)$ *if and only if* $q^{\mu} = 1 \pmod{k}$ *(see [DD99, 245]).*

Corollary 5.3. *Let* q *be a prime that is not divisible by* k *and* μ *a positive integer and* m *the first term of the arithmetic progression. Then the product*

$$\sum \theta^{\alpha\mu-\alpha_1} \sum \eta^{\beta\mu-\beta_1} \sum \omega^{\gamma\mu-\gamma_1} \sum \omega'^{\gamma'\mu-\gamma_1} \dots,$$

where the summation ranges over the respective $a, b, c, c', \dots$ *values of* $\theta, \eta, \omega, \omega', \dots$*, is non-zero with value* $\phi(k)$ *if and only if* $q^{\mu} = 1 \pmod{k}$ *(see [DD99, 247]).*

5.1.2 Analysis of Dedekind's use of characters and L-functions

Above we have described some of the most significant features of Dedekind's presentation of Dirichlet's theorem. Let us now consider what we can infer about his conception of characters, and what impact these features had upon his proof.

First, with respect to unification, we draw attention to the fact that Dedekind explicitly indicated that the characters were *functions*, and thus he presumably acknowledge that they were of the same sort as the functions studied in analysis. Second, with respect to reification, we see that Dedekind's approach was very similar to Dirichlet's, although there were some important differences.

Specifically, note the following:

[23] Note that Dedekind did not explicitly isolate these as theorems

1. The characters were introduced before Dedekind specialized his general *L*-functions to Dirichlet *L*-functions. Dedekind also introduced functional notation to represent the characters.

2. Dedekind was willing to attribute some properties to the characters, in particular complete multiplicity. But he was not willing to formulate some of the more substantial relations such as Corollary (3.13) in a similar way, and the characters were not made the subject of such theorems.

3. In a similar manner to Dirichlet, Dedekind classified the *L*-functions according to the roots of unity that were involved in their construction. That is to say, Dedekind relied on an *intensional* classification of the *L*-functions (and subsequently the characters).

4. Dedekind did not allow the characters themselves to occur in the range of a bound variable, although he did allow the *L*-functions themselves to play this role. He also allowed the summation to be over systems of roots of unity, rather than having to pick a representation of the characters and sum over the indices, as Dirichlet did.

5. Dedekind did not appear to take *L* as depending on the characters. For recall that he noted that *L* will behave differently "... depending on the roots of unity ..." [DD99, 240] appearing in their construction. Moreover, the specific character that an *L*-function corresponds to does not appear in the notation for *L* at all.

Let us say a little more about the above observations. The first three indicate that whilst Dedekind went some way towards introducing a new language to talk about characters without always having to refer to their construction, he did not fully make use of this. Indeed, as we have seen, he often reverted to talking about the components of the characters, instead of the characters themselves. Thus the answer to our clarificatory question is positive, although the answer to the second is only partially so: Dedekind did introduce the characters prior to constructing his specialized *L*-functions, and he did formulate complete multiplicity as a property that they enjoyed, but more complicated theorems were stated in terms of the components of the characters and the relations that held were not referred to as properties or relations of the characters themselves. The last three remarks above also suggest that the answers to our final three clarificatory questions is negative: Dedekind did not treat the characters as independent of their representations, he did not allow them to occur in the range of bound variables, and finally he did not conceive of *L* as dependent upon the characters.

Consequently, while Dedekind gave the characters an elevated status compared to Dirichlet by identifying and attributing at least some properties to them, he nonetheless treated them quite differently to other mathematical objects. Thus it is reasonable to assert that Dedekind did not treat the characters as legitimate mathematical objects, although his presentation was nonetheless a step towards that direction.

Let us now consider the impact that Dedekind's treatment of characters and L-functions had upon his proof. First of all, the organization of the proof aids its readability compared to Dirichlet's original presentation. For recall that in the introductory section, Dedekind isolated a general class of L-functions and argued that the Euler-product formula and certain series expansions held, provided the functions occurring in the L-series were completely multiplicative. In the next section, he introduced the characters, giving them their own definition, and noting that they are completely multiplicative. Thus, when we arrive at the point in the proof where we need these results for the Dirichlet functions, we do not have to divert our attention away from the overall structure of the proof and towards establishing a different lemma; we can instead recall from the introduction that the required theorems hold so long as, certain general conditions are met.

Furthermore, Dedekind's notation is far less cumbersome than Dirichlet's as he did not insist on including so much information about the construction of the characters in his notation for the L-functions. Thus, Dedekind was able to express relationships concerning the L-functions in a much more succinct manner. For example, recall that Dedekind could write

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

whereas Dirichlet wrote (changing his $1+\rho$ to s to make it consistent with Dedekind's symbols):

$$\sum \frac{1}{q^s}+\frac{1}{2}\sum \frac{1}{q^{2s}} + \frac{1}{3}\sum \frac{1}{q^{3s}} + \ldots$$
$$\frac{1}{K}\sum \Theta^{-\alpha_m \mathfrak{a}}\,\Phi^{-\beta_m \mathfrak{b}}\Omega^{-\gamma_m \mathfrak{c}}\Omega^{-\gamma'_{m'} \mathfrak{c}'} \ldots \log L_{\mathfrak{a},\mathfrak{b},\mathfrak{c},\mathfrak{c}'\ldots},$$

Thus we are less likely to get lost in Dedekind's notation than in Dirichlet's. And indeed, Dedekind's equation looks very similar to the modern variant (with the symbols appropriately changed to be consistent with Dedekind's):

$$\sum_{\chi \,\in\, U(\widehat{\mathbb{Z}/q\mathbb{Z}})} \overline{\chi(m)} \log L(s,\chi) = \phi(k) \sum_{q \equiv m \pmod{k}} \frac{1}{q^s} \;+\; O(1)$$

However, we should not be mislead by this similarity. For, as we drew attention to above, Dedekind's χ is not used, in this context, to denote a character. Rather it is used to denote the specific *value* expressed in modern terms as $\overline{\chi}(m)$. Thus, whilst Dedekind's notation is less overwhelming, it does not provide all of the salient information. Specifically, it does not indicate that there is an explicit relationship between the L-function and the value that its logarithm is multiplied by: namely the logarithm of the L function is multiplied by the value of the conjugate of the corresponding character when applied to the first term of the arithmetic progression. Thus Dedekind's choice of notation here obscures an important feature of the relationship.

Furthermore, Dedekind's reluctance to frame the orthogonality relations and their corollary in terms of the characters also impacts the effectiveness of his proof. Recall that the modern proof identifies the orthogonality relations as expressing a relation that holds between the characters modulo k and it emphasizes the strong parallelism between the two relations. However, note that the parallelism that the modern presentation makes so clear is lost in Dedekind's presentation: on the face of it, Theorems (5.1) and (5.2) appear to be quite unrelated. Moreover, it is not straightforward to parse what exactly it is that they are expressing and why they are important without the characters. With the characters, however, the relations become both simplified and clarified, and we gain a better perspective as to why they are central to the proof: they allow us to pick out primes in the residue class that we are interested in.

Consequently, whilst Dedekind's presentation of the theorem on primes in arithmetic progressions is appears to be better facilitate understanding more so than Dirichlet's original proof, there are nonetheless aspects of it that impede our ability to follow it, in particular:

1. Dedekind does not treat L as a function of the characters and thus the relationship expressed in the crucial equation $\phi(k)(\sum \frac{1}{q^s} + \frac{1}{2}\sum \frac{1}{q^{2s}} + \frac{1}{3}\sum \frac{1}{q^{3s}} + \ldots) = \sum \chi \log L$ is somewhat obscured.

2. Dedekind does not formulate the orthogonality relations and their corollary in terms of the characters, which makes them somewhat difficult to parse and obscures the important role that they play within the proof.

5.2 de la Vallée Poussin's presentation

We will now examine de la Vallée Poussin's presentation of Dirichlet's theorem. In particular, we will consider an article of his from 1895/6 entitled *Démonstration simplifiée du Théorèm de Dirichlet sur la progression arithmétique* and sections from his 1897 book *Recherches analytiques sur la théorie des nombres premiers*.

We will not provide an outline of his proofs, but we should note that they employ a different method than those used by Dirichlet and Dedekind, though characters and L-functions still take center stage. However, whilst Dirichlet and Dedekind were concerned with L functions that took a *real* variable s, de la Vallée Poussin allowed his L-functions to take *complex* variables.

With respect to notation, de la Vallée Poussin considered a general arithmetic progression of the form $Mx + N$ (assuming M and N to be coprime), and always assumed that p is an arbitrary prime number. In his earlier work, he focused solely on the case in which the common difference M is odd, whilst noting that the simplifications that his method had over Dirichlet's were "...completely independent ..." [24] [dlVP96, 11] of the parity of M. In his *Recherches analytiques*, however, his work on the characters and L-functions was built up from the simplest case ($M = p$) to the most complicated (M is an arbitrary natural number) through two intermediate cases: $M = p^{\alpha}$ for some $\alpha \in \mathbb{N}$ and $M = 2^c$ for some $c > 1$. For the sake of simplicity, we will focus on his presentation in the 1895/6 article, referring to the later work as necessary.

In his 1895/6 paper, de la Vallée Poussin introduced the characters modulo M in a similar manner to the constructions used by Dirichlet and Dedekind. And, just like Dedekind, de la Vallée Poussin indicated that the characters were to be conceived of as functions. Indeed, in a preliminary section prior to defining the characters, he presented some general considerations about the series

$$f(s) = \frac{\chi(1)}{1} + \frac{\chi(2)}{2^s} + \ldots \frac{\chi(n)}{n^s} + \ldots$$

Of the $\chi(n)$, de la Vallée Poussin made the following remark [25]:

> The *functions* $\chi(n)$ are real or complex, they are linked to a certain integer M and enjoy the two characteristic properties expressed in the two equalities:
>
> $$\chi(1) + \chi(2) + \ldots + \chi(M) = 0,$$
>
> $$\chi(m) = \chi(n), \text{ if } m \equiv n \pmod{M}$$
>
> [dlVP96, 8, emphasis added].

[24] "...complètement indépendentes..."

[25] Les fonctions $\chi(n)$ sont réelles ou complexes, elles sont liées à un certain nombre entier M et jouissent de deux propriétes caractéristiques exprimeés dans la deux égalités

$$\chi(1) + \chi(2) + \ldots + \chi(M) = 0,$$

$$\chi(m) = \chi(n), \text{ if } m \equiv n \pmod{M}$$

As de la Vallée Poussin defined the characters in the very next section and established that they satisfied the conditions quoted above, he undoubtedly conceived of them as functions.

However, whilst Dedekind's characters were implicitly defined for all natural numbers, de la Vallée Poussin, like Dirichlet, restricted the domain of the characters modulo M to natural numbers coprime to M. Assuming that $M = p_1^{\alpha_1} p_2^{\alpha_2} \dots$ with the p_i odd primes, and putting $\pi_1 = \phi(p_1^{\alpha_1}), \pi_2 = \phi(p_2^{\alpha_2}), \dots$, de la Vallée Poussin defined a character of n as

$$\chi(n) = \omega_1^{\nu_1} \omega_1^{\nu_1} \dots .$$

He further explained "One forms the different characters [from the above] by substituting the different roots of the equations $[\omega_1^{\pi_1} = 1, \omega_2^{\pi_2} = 1, \dots]$[26]" [dlVP96, 12]. Noting that there were $\phi(M) = \pi_1 \pi_2 \dots$ such characters, de la Vallée Poussin distinguished between them when necessary by indexing them via a subscript, so that, for modulus M, the characters are $\chi_1, \chi_2, \dots, \chi_{\phi(M)}$.

Whilst the proofs given by Dirichlet, Dedekind and modern writers divide the characters into three classes at the beginning of the proof, de la Vallée Poussin did something different in his 1895/6 paper. Specifically, he distinguished only between the principal character and the non-principal characters at the start of the proof[27]:

> One gives the name *principal character* to that which corresponds to roots which are all equal to +1
>
> $$\omega_1 = 1, \omega_2 = 1, \dots$$
>
> The principal character is therefore always equal to unity and it plays a special role in the considerations that follow.
>
> The other characters will sometimes be real and sometimes imaginary, but their modulus is always equal to unity [dlVP96, 13]

[26]On formera des caractères différents en remplaçant les unes par les autres les différents racines des équations.

[27]On donne le nom de *caractère principal* à celui qui correspond aux racines toutes égales à +1

$$\omega_1 = 1, \omega_2 = 1, \dots$$

Le caractère principal est donc toujours égal à l'unité et il joue un rôle special dans les considérations qui vont suivre.

Les autres caractères seront tantôt réeles et tantôt imaginaires, mais leur module est toujours égal à l'unité.

However, the distinction between real and imaginary characters is nonetheless implicit within his presentation and becomes clear as his proof develops. Indeed, he needed to prove that the L-function associated with χ does not vanish when χ is a non-principal character. To demonstrate this, he first established that such an L-function can vanish only for at most one (non-principal) character [dlVP96, 20-21], before proceeding to show that if it vanishes, then the associated character must take the values of ± 1 for all values of n coprime to M [dlVP96, 24]. He then completed the proof by establishing that such a series corresponding to a character that takes only real values cannot vanish [dlVP96, 24-29]. Thus although we do not see him explicitly make a distinction between the characters at the start of his proof, it is nonetheless teased out within the proof, since he isolated exactly the real, non-principal characters. The complex characters are then those non-principal characters that are left over. Moreover, as the characterization of real characters is done in terms of its *values*, it is extensional, in contrast to both Dirichlet and Dedekind.

The approach taken in the 1897 book was, however, quite different. Indeed, here de la Vallée Poussin made the distinction for characters of a prime modulus, a modulus that is a power of a prime, and a modulus that is a power of 2. Thus, for example, for the simplest case of a prime modulus, de la Vallée Poussin wrote [28]:

> One calls the *principal character* that which correspond to the root +1: it is equal to unity for all the 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 according to the number n.
>
> We give to all of the other characters the name of imaginary characters, though they may have a real value for certain particular numbers. Their modulus is always equal to unity [dlVP97, 19]

While he did not appear to explicitly formulate the distinction for characters of an arbitrary modulus, it still carried over to them, since he utilized the same terminology (see e.g. [dlVP97, 67]) and presumably conceived of real and imaginary such characters in the same way.

Returning to his 1896 presentation, immediately after introducing the characters, de la Vallée Poussin formulated four key results[29]:

[28]One 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 *caractères imaginaires*, quoiqu'ils puissent avoir une valeur réelle pour certains nombres particuliers. Leur module est toujours égale a l'unité.

[29]

1. Whatever characters we consider, in other words, whatever the system of roots of equations $[\omega_1^{\pi_1} = 1, \omega_2^{\pi_2} = 1, \ldots]$ that one chooses, one has for any two numbers n, n' coprime to M the fundamental relation

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

...

2. Since two numbers which are congruent with respect to the modulus M have the same index, one has, for any character

$$\chi(n) = \chi(n') \text{ if } n \equiv n' \pmod{M} \ldots$$

---

1. Quel que soit le caractère que l'on considère, en d'autres termes, le système de racines des équations $[\omega_1^{\pi_1} = 1, \omega_2^{\pi_2} = 1, \ldots]$ que l'on choisisse, on aura pour deux nombres quelconques n et n' premiers avec M la relation fonctionnelle

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

...

2. Comme deux nombres qui sont congruents suivant le module M ont même indicateur, on a, pour un caractère quelconque

$$\chi(n) = \chi(n') \text{ si } n \equiv n' \pmod{M} \ldots$$

3. Pour chacun des caractères, à l'exception du caractère principal, la somme étendue à tous les nombres premiers avec M et inférieurs à M s'annule, c'est-à-dire qu'on a

$$\sum_n \chi(n) = 0;$$

dans le cas du caractère principal, on a, au contraire;

$$\sum_n \chi(n) = \phi(M) \ldots$$

4. Considérons, d'autre part, la somme étendue à tous les caractères, c'est-à-dire à tous les systèmes de racines

$$\mathcal{S}_\chi \chi(n) = \mathcal{S}_\omega \omega_1^{\nu_1} \omega_2^{\nu_2} \ldots,$$

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

$$\mathcal{S}_\chi \chi(n) = 0,$$

à la seule exception près du cas où

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

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

$$\mathcal{S}_\chi \chi(n) = \phi(M).$$

3. . . . For each of the characters, with the exception of the principal character, the sum extending over all the numbers coprime with M and less than M vanishes, that is to say that one has

$$\sum_n \chi(n) = 0;$$

in the case of the principal character, one has, on the contrary,

$$\sum_n \chi(n) = \phi(M) \ldots$$

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

$$\mathcal{S}_\chi \chi(n) = \mathcal{S}_\omega \omega_1^{\nu_1} \omega_2^{\nu_2} \ldots,$$

. . . For all numbers n, the sum extending over the totality of characters

$$\mathcal{S}_\chi \chi(n) = 0,$$

with the sole exception the cases where $n \equiv 1 \pmod{M}$, because then all the indices are zero and one has

$$\mathcal{S}_\chi \chi(n) = \phi(M).$$

[dlVP96, 13-15]

Similar results were formulated in his *Recherches analytiques* for each class of characters he introduced (see e.g. [dlVP97, 20], [dlVP97, 40], [dlVP97, 44], [dlVP97, 48]). Thus, we see that de la Vallée Poussin undertook a more complete study of the characters themselves than either Dirichlet or Dedekind. For recall that Dirichlet did not attribute any properties to them whatsoever, and Dedekind only formulated complete multiplicity as a property of the characters. In particular, Dedekind did not formulate the orthogonality relations as holding of the characters, but of their components, whereas we see in (3) and (4) above that de la Vallée Poussin formulated these relations as relations between the characters themselves.

de la Vallée Poussin's formulation of the second orthogonality relation in (4) above is of particular importance. For he introduced a new symbol, $\mathcal{S}$, to denote a sum that ranges over the *characters*, using the bound variable χ to denote these characters. He did specify what the sum over characters is in terms of the components of the characters, but throughout the proof he felt comfortable enough to continue to use the $\mathcal{S}$ symbol without constantly having to refer to its interpretation in terms of the construction data.

However, whilst de la Vallée Poussin's results are formulated as being about the characters themselves, we should note that the proofs are nonetheless couched in terms of their components. Thus, for example, his proof of the orthogonality relations relies on explicit calculations concerning the components. As an example, we will present his proof of the first orthogonality relation, result 3, listed above[30]:

> The second member $[\sum_\nu \omega_1^{\nu_1}\omega_2^{\nu_2}\ldots]$, in which the sum extends over all systems of possible exponents, is equal to the product
>
> $$\begin{array}{l}(1+\omega_1+\omega_1^2+\ldots+\omega_1^{\pi_1-1})\\(1+\omega_2+\omega_2^2+\ldots+\omega_2^{\pi_2-1})\\\cdots\cdots\cdots\end{array}$$
>
> If not all of the roots $\omega_1, \omega_2, \ldots$ are equal to $+1$, there will be at least one factor of this product which vanishes, and we can announce the following theorem [Result 3, above] [dlVP96, 14]

This is in contrast to the modern presentation. For whilst we do utilize facts about the components of the characters in establishing that $G \cong \widehat{G}$ and that if $g \in G, g \neq 1_G$ then there exists $\chi \in \widehat{G}$ such that $\chi(g) \neq 1$, the proofs of the orthogonality relations do not make explicit use of the components [31].

Having introduced the characters, de la Vallée Poussin then introduced the L-functions. In his 1896 article, however, he did not utilize a specific functional abbreviation for the L-functions. Instead, he simply refereed to their sum or product form. In his 1897 work, however, de la Vallée Poussin did utilize a functional abbreviation. He introduced L-functions first for a particular type of characters of modulus M called *proper characters* [32], and then used these to obtain the properties of L-functions corresponding to other kinds of characters. Thus, for the proper

[30]Le second membre $[\sum_\nu \omega_1^{\nu_1}\omega_2^{\nu_2}\ldots]$, dans lequel la somme s'étend à tous les systèmes d'exposants possibles, est égal au produit

$$\begin{array}{l}(1+\omega_1+\omega_1^2+\ldots+\omega_1^{\pi_1-1})\\(1+\omega_2+\omega_2^2+\ldots+\omega_2^{\pi_2-1})\\\cdots\cdots\cdots\end{array}$$

Si toutes les racines $\omega_1, \omega_2, \ldots$ ne sont pas égales à $+1$, il y aura un facteur au moins de ce produit qui s'annulera, et nous pouvons énoncer le théorème suivant [Result 3, above]

[31]They do implicitly rely on calculations with the components, however, since they are established by using the isomorphism result and its consequence just mentioned.

[32]Roughly speaking, these are characters which cannot be obtained for modulus M', where $M' < M$.

characters, he introduced the corresponding L-function as follows[33]:

> We define the function $Z(s,\chi)$, for $\mathbb{R}(s) > 1$, by the absolutely convergent expressions:
>
> 1. $Z(s,\chi) = \sum_{n=1}^{\infty\prime} \frac{\chi(n)}{n^s} = \prod \left(1 - \frac{\chi(q)}{q^s}\right)^{-1}$
>
> where n designates successively all the integers prime to M and q all the prime numbers not dividing M. [dlVP97, 56]

Thus de la Vallée Poussin denoted the L-functions by $Z(s,\chi)$. Note that unlike Dirichlet or Dedekind, but like the modern authors, he allowed the character χ to occur in the argument position of the L-function.

However, whilst de la Vallée Poussin's proof is quite modern in this respect, we should remark that when dealing with central equations involving the L-functions, he adopted an attitude that is strikingly similar to Dirichlet's: he appeared reticent to *quantify* over the characters. Indeed, in his 1897 work, he wrote [34]:

> ...one finds the *fundamental equation*
>
> $$(E)\ldots -\lim_{s=1}(s-1)\frac{Z'(s,\chi)}{Z(s,\chi)} = \lim_{s=1}(s-1)\sum_q \chi(q)\frac{lq}{q^s},$$
>
> and this equation represents in reality $\phi(M)$ distinct ones, which result from exchanging the characters among themselves. [dlVP97, 65].

Thus, whilst we saw above that de la Vallée Poussin was willing to allow characters to appear in the range of a summation index, like Dirichlet he still seemed to be wary of certain forms of quantification.

Moreover, unlike modern authors, de la Vallée Poussin did not allow the χ occurring in the argument position of the L-function to occur in the range of a summation symbol. For recall that in the modern presentation, we utilized expressions

[33]Nous définirons la fonction $Z(s,\chi \pmod{M})$, pour $\Re(s) > 1$ par les expressions absolument convergentes

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

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

[34]...on trouve l'*équation fondamentale*

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

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

such as:

$$\sum_{\chi \in \widehat{U(\mathbb{Z}/q\mathbb{Z})}} \overline{\chi(a)} \log L(s,\chi) = \phi(q) \sum_{p \equiv a \pmod q} \frac{1}{q^s} + O(1) \tag{11}$$

Thus we quantified over the χ appearing in the argument position of the L-function, and thus emphasized that the L-function is to be considered as a function of χ in this context. Thus we might suspect that whilst de la Vallée Poussin put the characters in the argument position of the L-functions, he did not properly conceive of them as functions of the characters. However, whilst he did not do this in his proof of Dirichlet's theorem, when dealing with a generalized notion of characters in his *Recherches analytiques*, he allowed generalized characters appearing in the argument of another function to appear in the range of a bound variable. Without giving the details, he obtained the following equation, where χ is a character and $L(s,k,\chi)$ is explicitly called a function [dlVP97, 5th part, 20][35]:

$$\phi(M) \lim_{s=1}(s-1) \sum_q [k(c_q) + k(c_q^{-1})] \frac{lq_1}{q_1^s} = \lim_{s-1} \mathcal{S}_\chi \frac{L'(s,k,\chi)}{L(s,k,\chi)}. \tag{12}$$

This indicates that de la Vallée Poussin did allow for certain functions to be conceived of functions of characters, under certain circumstances, though not within the specific context of Dirichlet's theorem.

5.2.1 Analysis of de la Vallée Poussin's use of characters and L-functions

We have highlighted above some of the most significant features of de la Vallée Poussin's treatment of characters and L-functions. Thus we now come to consider what these features reveal about his conception of characters and L-functions, and how this impacts upon his presentation of Dirichlet's theorem.

We first make a brief remark concerning the topic of unification. As mentioned above, in his 1895/6 work, de la Vallée Poussin indicated that he conceived of the characters as functions. Moreover, as we have seen, he used functional notation to abbreviate them. Additionally, we have seen that there are indications that he also allowed for functions of characters, or, more generally functions of functions. Thus de la Vallée Poussin appeared to be working with a more unified conception of function than that which Dirichlet or Dedekind displayed in their presentations of the proof of Dirichlet's theorem. For not only was he willing, like Dedekind, to

[35] $L(s,k,\chi)$ is built up from several components, and its definition is given on page 13 of the fifth part of the *Recherches analytiques*. It is also explicitly called a function on this page. That χ is a character can be seen by looking at the definition of one of the components of $L(s,k,\chi)$ on page 9 in the fifth part.

allow for functions defined on subsets of the natural numbers, he was also willing to allow for functions of other functions.

Let us now focus on the issue of reification. We have noted above that de la Vallée Poussin's approach was in many ways much more like the modern presentation than Dirichlet's or Dedekind's. Specifically, we draw attention to the following points:

1. de la Vallée Poussin, like Dedekind and modern authors, introduced functional notation to abbreviate the characters.

2. He subjected the characters to a much more detailed study in their own right than either Dirichlet or Dedekind. In particular, note that he formulated the orthogonality relations as a relation that held of the *characters* themselves, and did not reduce them to statements about the components of the characters as Dedekind had done. However, his proofs of the orthogonality relations relied on explicit calculations about the components of the characters, whereas the modern proofs only implicitly rely on information about their construction via the theorem that $G \cong \widehat{G}$ and its immediate consequences.

3. de la Vallée Poussin's classification of the characters had both intensional and extensional features. Note that his definition of the principal character in both his 1896 article and 1897 book was intensional, since it was defined in terms of the roots involved in its construction. However, in both works, he clarified this definition in extensional terms, remarking that it was always equal to 1. His definition of real characters modulo p in his 1897 work similarly contains a mix of intensional and extensional features. However, the classification of real characters that was teased out of his 1895/6 proof was purely extensional, defining such characters in terms of the values they can take.

4. Unlike Dirichlet or Dedekind, de la Vallée Poussin did allow his characters to appear in the range of a summation index. And, we see that he introduced a new variable binding operator that applied specifically to the characters $\mathcal{S}_\chi$. However, he was also reticent about quantifying over them in certain other contexts.

5. de la Vallée Poussin allowed for dependencies upon the characters. Specifically, he introduced the notation $Z(s,\chi)$ for the L-functions and in later sections of his 1897 work, allowed the χ in the argument position of $L(s,k,\chi)$ to appear in the range of a bound variable, thus emphasizing its functional nature.

The above points indicate that de la Vallée Poussin introduced a new language to allow him to talk about the characters without having to refer back to their components. However, he did not entirely avoid falling back into talk of their components, since his proofs of results (1)-(4) still relied explicitly upon calculations with construction data, and his classification of the characters was in part intensional. The fact that he still appeared uncomfortable quantifying over the characters in some contexts and that he did not use the symbol $\sum$ when summing over the characters also indicates that he did not quite view the characters as having the same status as natural numbers or other mathematical objects. However, the fact that he introduced a new symbol to allow the characters to appear in the range of a bound variable still indicates that he was willing to work with the characters in ways that paralleled his treatment of other mathematical entities, and to a much greater extent than Dirichlet or Dedekind.

Consequently, we see that the answers to all but our third clarificatory questions are affirmative: the characters are given their own definition, they are studied in their own right, they can appear in the range of some bound variables, and there can be dependencies on them. The third question, "Is the treatment of the candidate objects independent of their different representations?", receives a partially positive and partially negative answer: for as we have seen, de la Vallée Poussin treats his characters both intensionally and extensionally.

The above considerations thus strongly suggest that with de la Vallée Poussin we see the characters being treated very much like legitimate mathematical objects, although not yet on a par with the natural numbers. Let us now consider how this impacted his presentation of the proof.

First, we should note that the study of the characters prior to the main proof of Dirichlet's theorem significantly aids the intelligibility of de la Vallée Poussin's presentation, even compared to Dedekind's proof. For, as we have seen, de la Vallée Poussin formulated four important features of the characters, in particular the orthogonality relations, unlike Dedekind who only formulated their complete multiplicity. Thus at stages in the proof in which we need to appeal to the orthogonality relations, we do not have to stop and consider how to establish that these relations hold: we can simply recall what was previously proved.

Moreover, we should note that de la Vallée Poussin's presentation of the orthogonality relations helps improve the clarity of his proof. Indeed, since he allowed the characters to appear in the range of a bound variable, he was able to express them succinctly, in a manner which highlighted the duality between them and which made it clear that they allow us to "pick out" primes in the residue class that we are interested in.

de la Vallée Poussin's notation for the L-functions, namely $Z(s,\chi)$, also improves the effectiveness of his proof. Unlike Dirichlet, his notation does not provide information about the construction of the characters, and thus his notation is clearer and less clumsy. But, unlike Dedekind, his notation allows for a more fine grained distinction between the L-functions. For recall that Dedekind's notation only allowed us to distinguish between L-functions that were in different classes, or between an L-function and its conjugate if it corresponded to a complex character. de la Vallée Poussin's notation, however, allows us to distinguish between each of the L-functions by distinguishing the characters that they correspond to.

However, we should also note that not all of the features of de la Vallée Poussin's proof appear to be desirable. In particular, his use of a different symbol for summation over the characters (and certain collections of complex numbers) compared to summation over the natural numbers could become problematic. For if, as appears to be the case, de la Vallée Poussin utilized a different symbol to signify summation over different types of mathematical "things", then we can imagine situations arising when we need to use numerous different types of summation symbols. But then we will have to keep track of what each summation symbol means and what it sums over, thus increasing the cognitive effort required to understand and follow the proof. This scenario does not happen in de la Vallée Poussin's proof, but nonetheless it appears to be a plausible situation if we utilize different summation symbols for different types of summands.

Consequently, de la Vallée Poussin's presentation has many features which aid the reader's understanding of the proof, specifically his detailed study of the characters in their own right, his clear presentation of the orthogonality relations, and his notation for the L-functions. However, we also noted that treating the summation over characters and other mathematical "entities" differently to summation over the natural numbers could cause difficulties if we want to sum over various different kinds of things.

5.3 Hadamard's treatment of Dirichlet characters and L-functions

Let us now come to consider Hadamard's treatment of Dirichlet characters and L-functions. In a famous paper of 1896, Hadamard proved an important result in number theory: the *Prime Number Theorem*, which de la Vallée Poussin proved independently in the same year in his *Recherches analytiques*. This result states that the number of primes less than or equal to x, denoted $\pi(x)$, is asymptotic to $\frac{x}{\ln(x)}$, i.e. that $\frac{\pi(x)\ln(x)}{x}$ approaches 1 as x approaches infinity. Furthermore, in the same paper, Hadamard proved a corresponding result about primes in arithmetic progressions, dubbed the Prime Number Theorem for Arithmetic Progressions: "...the sum of logarithms of prime numbers less than or equal to x and which are contained in a determined arithmetic progression of common difference k is asymptotic to

$\frac{x}{\phi(k)}$"[36] [Had96, 219]. Thus in his paper, Hadamard utilized important results concerning L-functions and characters that appear in the proof of Dirichlet's theorems on primes in arithmetic progressions. Interestingly, he remarked in a footnote that "We conform to the notations employed in the *Vorlesungen über Zahlentheorie*, edited by Dedekind, 1863 edition, supplement VI" [37] [Had96, 205]. However, as we shall see, there were some differences between Hadamard's use of notation for the characters and L-functions compared to Dedekind's. In what follows, we will let M be the first term of an arithmetic progression, k the common difference, again assuming that M and k are coprime, and q will denote a prime number.

As with each of the previous mathematicians we have considered, Hadamard's introduction of the characters involved a construction of roots of unity and primitive roots of a given modulus. However, he explicitly defined characters as functions on the entire natural numbers by making use of a *definition by cases*. As we have seen Dirichlet and de la Vallée Poussin only defined the characters for certain subsets of the natural numbers, and while Dedekind implied that they could be extended so as to be defined for all natural numbers, he did not explicitly do so. In particular, Hadamard wrote the following when defining the characters:

> Dirichlet introduced the function
>
> $$\psi_v(n) = \begin{cases} 0, & \text{if } n \text{ is not coprime to } k \\ \theta^{\alpha}\eta^{\beta}\omega^{\gamma}\omega'^{\gamma'}\ldots, & \text{if } n \text{ is coprime to } k \end{cases}$$
>
> $\alpha, \beta, \gamma, \gamma', \ldots$ being the indices[38] of n [the index v is to distinguish between the one and the other of the $\phi(k)$ functions ψ corresponding to the different possible choices of the roots $\theta, \eta, \omega, \omega', \ldots$][39] [Had96, 207]

Hadamard then introduced the L-functions. Like de la Vallée Poussin, his L-functions took a *complex* variable. He described Dirichlet's introduction of the

[36]... la somme des logarithmes des nombres premiers inférieurs à x et compris dans une progression arithmétique déterminée de raison k est asymptotique à $\frac{x}{\phi(k)}$.

[37]Nous nous conformons aux notations aux notations employées dans les *Vorlesungen über Zahlentheorie*, éditées par Dedekind, édition de 1863, supplément VI.

[38]$\theta, \eta, \omega, \omega', \ldots$ are, as previously, roots of unity corresponding to the prime decomposition of k.

[39]Dirichlet introduit la fonction

$$\psi_v(n) = \begin{cases} 0, & \text{if } n \text{ is not coprime to } k \\ \theta^{\alpha}\eta^{\beta}\omega^{\gamma}\omega'^{\gamma'}\ldots, & \text{if } n \text{ is coprime to } k \end{cases}$$

$\alpha, \beta, \gamma, \gamma', \ldots$ étant les indices de n [l'indice v a pour but de distinguer les unes des autres les $\phi(k)$ fonctions ψ correspondant aux différents choix possibles des racines $\theta, \eta, \omega, \omega', \ldots$]

L-functions as follows, where $\xi_r(s) := \frac{i}{2\pi}\Gamma(1-s)\int(-x)^{s-1}\frac{e^{(k-r)x}}{e^{kx}-1}dx$ and $\Gamma(x)$ is the Gamma function:

> He then forms the series (periodic in the sense indicated above)
>
> $$(14)\ \ L_\nu(s) = \sum_{n=1}^{\infty}\frac{\psi_\nu(n)}{n^s} = \sum_{r=1}^{k}\xi_r(s)\psi_\nu(r)\ \ [\nu = 1, 2, \ldots, \phi(k)]$$
>
> equal to the infinite product
>
> $$(15)\ \ L_\nu(s) = \prod \frac{1}{1-\frac{\psi_\nu(q)}{q^s}}$$
>
> in which q must be replaced successively by all the prime numbers [Had96, 207][40]

Thus Hadamard's notation for the L-functions is different from the notation used by Dirichlet, Dedekind and de la Vallée Poussin. Note first that the index ν is not part of the construction data associated with the character, and so Hadamard's notation is considerably different from Dirichlet's original. And as each character receives its own index, the notation allows us to distinguish between any two L-functions, whereas Dedekind could only distinguish between L-functions within different classes or between an L-function and its conjugate. Finally, de la Vallée Poussin allowed the characters to appear within the argument position, but Hadamard did not go this far. Indeed, not only did Hadamard utilize a number instead of a character, this index was placed as a subscript, and not in the argument position.

As with previous mathematicians, Hadamard also classified the L-functions into three classes:

> The series L_ν fall into three categories: the first contains only one series L_1, which corresponds to
>
> $$\theta = \eta = \omega = \omega' = \ldots = 1;$$

[40] Il forme ensuite la série (périodique au sens indiqué ci-desus)

$$(14)\ \ L_\nu(s) = \sum_{n=1}^{\infty}\frac{\psi_\nu(n)}{n^s} = \sum_{r=1}^{k}\xi_r(s)\psi_\nu(r)[\nu = 1, 2, \ldots, \phi(k)]$$

égal au produit infini

$$(15)\ \ L_\nu(s) = \prod \frac{1}{1-\frac{\psi_\nu(q)}{q^s}}$$

dans lequel q doit être remplacée successivement par tous les nombres premiers.

the second contains all the series L, for which the numbers $\theta, \eta, \omega, \omega', \ldots$ are equal to +1 or to -1 (with the exception of L_1); the third, the series corresponding to the case where at least one of the numbers is imaginary. These last are conjugated two by two ...[41] [Had96, 207]

Thus, like Dirichlet and Dedekind, Hadamard's classification of the L-functions and characters is intensional: it relies on information about how the characters are constructed.

Hadamard utilized his notation for the characters and L-functions to provide a succinct way of expressing fundamental relations, such as Equation (2), above. Recall that in modern notation (slightly altering the choice of variables to make it consistent with Hadamard's presentation), this is expressed as:

$$(16) \qquad \sum_{\chi \in U(\widehat{\mathbb{Z}/k\mathbb{Z}})} \overline{\chi(m)} \log L(s,\chi) = \phi(k) \sum_{q \equiv k \pmod k} \frac{1}{q^s} + O(1).$$

Hadamard, however, wrote the following:

> The fundamental equation utilized by Dirichlet for the demonstration of his theorem is
>
> $$\sum_{v} \frac{\log L_v(s)}{\psi_v(m)} = \phi(k)\left(\sum \frac{1}{q^s} + \frac{1}{2}\sum{}' \frac{1}{q^{2s}} + \frac{1}{3}\sum{}'' \frac{1}{q^{3s}} + \ldots\right),$$
>
> where m is some integer prime to k and where the signs $\sum, \sum', \sum'', \ldots$ extend, the first over all prime numbers q such that $q \equiv m \pmod k$, the second over prime numbers q such that $q^2 \equiv m \pmod k$, etc [42] [Had96, 209]

[41] Les séries L_v se rêpartissent en trois catégories: la première comprend une seule série L_1, celle qui correspond à

$$\theta = \eta = \omega = \omega' = \ldots = 1;$$

la seconde comprend toutes les séries L, pour lesquelles les nombres $\theta, \eta, \omega, \omega', \ldots$ sont égaux à +1 ou à −1 (à l'exception de L_1); la troisième, les séries correspondant aux cas où l'un au moins de ces nombres est imaginaire. Ces dernières sont conjugées deux à deux...

[42] L'équation fondamental utilisée par Dirichlet pour la démonstration de son théorème, est

$$(16) \qquad \sum_{v} \frac{\log L_v(s)}{\psi_v(m)} = \phi(k)\left(\sum \frac{1}{q^s} + \frac{1}{2}\sum{}' \frac{1}{q^{2s}} + \frac{1}{3}\sum{}'' \frac{1}{q^{3s}} + \ldots\right)$$

où m est un entier quelconque premier avec k et où les signes $\sum, \sum', \sum'', \ldots$ 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.

Thus Hadamard allowed the indices of the characters and the corresponding indices of the *L*-functions to appear in the range of a bound variable. And although Hadamard, like Dirichlet and Dedekind, had his summation indices ranging over particular numbers, there are some important differences we should note. For recall that the index Hadamard associated with the character provides us with no information about how it is constructed. It is not part of the construction data, and is just a number that we use to identify and distinguish between the characters. For Dirichlet and Dedekind, however, the situation is quite different. Indeed, for Dirichlet the construction data just *is* what the characters are, and talk about characters is to be expressed as talk about their construction data. And this is also true, though to a lesser extent, of Dedekind. Thus Hadamard appears to be between Dirichlet and Dedekind, on one hand, and de la Vallée Poussin, on the other. For he did not feel the need to reduce talk about summation over characters in terms of their components, but yet still did not feel comfortable with allowing the summations to range over the characters themselves.

5.3.1 Analysis of Hadamard's use of characters and *L*-functions

Above we have highlighted some of the important features of Hadamard's presentation of Dirichlet characters and *L*-functions. Let us now consider what this reveals about his conception of them and how it impacts upon his presentation. First, we make a brief remark concerning the unification of the function concept. Whilst Hadamard was indeed comfortable with treating the characters as functions, he did not appear to be willing to go as far as de la Vallée Poussin in treating the *L*-functions as functions of the characters. Indeed, he did not feel comfortable enough to allow the *character* to occur in the argument position of the *L*-function. Thus de la Vallée Poussin appeared to be comfortable working with a more general, unified conception of function than Hadamard did.

Let us now consider the issue of the reification of the function concept. First, we should note that Hadamard did not present a thorough study of the characters in their own right, though presumably this is because the results concerning primes in arithmetic progressions were not his main concern in his 1896 paper. Additionally, we isolated the following important points about Hadamard's treatment of the characters and *L*-functions in the previous section:

1. Hadamard's presentation of the characters and *L*-functions meant that they were still dependent upon their representation.

2. The characters themselves did not appear within the range of bound variables, but the number used to index them did.

3. Hadamard did not explicitly identify the L-functions as functions or dependencies upon the characters. However, his notation for the L-functions referred to the index for the character and the fact that he was willing to allow these to appear in the range of the summation indices suggests that Hadamard may have viewed the L-functions as depending on these indices.

Thus, in terms of language, we see that whilst Hadamard introduced appropriate new notation for the characters, he was still reluctant to operate on the characters in certain respects. Consequently, whilst we have a positive answer to our first clarificatory question, "are the characters given their own definition?", we see that the answers to the remaining clarificatory questions are negative. However, despite these negative answers, Hadamard nonetheless appears to be between Dirichlet and Dedekind, on the one hand, and de la Vallée Poussin, on the other regarding his treatment of the characters and L-functions.

Let us now consider how Hadamard's treatment of the characters and L-functions impacted upon his presentation of the results. First, note that Hadamard's use of indices enumerating the characters provides us with a succinct way of referring to the characters in our notation and to distinguish between them, without having to keep all of the construction data in mind. And this notation provides a short and precise way of expressing fundamental relations between the characters and L-functions, since we can simply have summation indices ranging over the index of the characters.

However, we should also note that Hadamard's treatment of the characters and L-functions has some disadvantages. First of all, recall that in Hadamard's notation, the index associated with the character does not appear in the argument position of the L-function, but only as a subscript. This makes it less prominent than the complex variable s and also seems to indicate that the index is not given the same status as the complex variable. And, if we are dealing with functions that are associated to a collection of different variables or other functions, then how are we to proceed? That is, for example, if we have a function of s that is associated with a character modulo k and a character modulo k', then how should we represent this? Should both of the different characters appear as a subscript? Should one appear as a subscript and one appear as a superscript? And what happens if we then need to consider a function of complex variable s that corresponds to three characters modulo k, k', and k''? If the index of the characters appeared in the argument position, then these awkward questions would not arise. For we would have a uniform way to denote the corresponding function: in each case we would introduce a new argument position and place a new variable in that position for each of the elements that the function corresponds to.

Furthermore, we see that such scenarios highlight an additional difficulty: it seems that, potentially, relying on the use of such indices could increase the cognitive burden the reader must bear. In the case where one type of index is being used there is perhaps little extra cognitive work, but as we have to deal with more, we have to remember what each index is standing in for. Working directly with what the corresponding character instead of the index associated with it would alleviate this problem.

Consequently, we see that Hadamard's presentation and the use of indices to stand in for the characters places him somewhere between Dirichlet and Dedekind, on the one hand, and de la Vallée Poussin on the other. And, whilst his use of these indices allowed for succinct notation for the *L*-functions and relations between the *L*-functions and characters, it could nevertheless create more cognitive burden for the reader in situations where more than one kind of such index is required.

5.4 Landau's treatment of Dirichlet characters and *L*-functions

Landau is the final mathematician whose work we will consider. He presented proofs of Dirichlet's theorem on arithmetic progressions in his 1909 work *Handbuch der Lehre von der Verteilung der Primzahlen* and his 1927 work *Vorlesungen über Zahlentheorie*, although the approach taken in each of these is different in certain respects. In particular, his treatment of characters and *L*-functions in his *Primzahlen* is, in some ways, similar to Hadamard's, whereas his *Vorlesungen* is much like the modern proof. We begin with an examination of his *Primzahlen* before discussing his *Vorlesungen*.

5.5 Landau's *Primzahlen*

In what follows, we will adhere to Landau's usage of l for the first term of an arithmetic progression, k for the common difference (where l and k are assumed, as usual, to be coprime) and p for a prime.

In his 1909 work, Landau introduced the characters in the same way as his predecessors: by utilizing the explicit construction in terms of roots of unity and primitive roots. He introduced them as follows [Lan09, 402][43]:

> Then the definition of $\chi_{(a_1,a_2,\ldots,a_r,a,b)}(n)$ runs as follows:
>
> 1. For all n which have a common divisor with k,
>
> $$\chi_{(a_1,a_2,\ldots,a_r,a,b)}(n) = 0$$
>
> 2. If n is coprime to k and has the index system
>
> $$\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha, \beta,$$
>
> then
>
> $$\chi_{(a_1,a_2,\ldots,a_r,a,b)}(n) = \rho_1^{a_1\alpha_1}\rho_2^{a_2\alpha_2}\ldots\rho_r^{a_r\alpha_r}\rho_{r+1}^{a\alpha}\rho_{r+2}^{b\beta}$$

The index system $\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha, \beta$ is the collection of indices of n with respect to the primitive roots corresponding to the components of the prime decomposition of k, the ρ_i are primitive roots of unity corresponding to these components, and the $a_1, a_2, \ldots, a_r, a, b$ serve to distinguish the characters, just like Dirichlet's $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{c}', \ldots$. Thus here we see that Landau included much of the construction data in the notation for the characters, in a manner similar to Dirichlet.

However, after proving that there are $h = \phi(k)$ such functions, Landau abbreviated his notation (see [Lan09, 404])[44]:

---

[43] Dann lautet die Definition von $\chi_{(a_1,a_2,\ldots,a_r,a,b)}(n)$:

1. Für alle n, die mit k einen gemeinsamen Teiler besitzen, ist

$$\chi_{(a_1,a_2,\ldots,a_r,a,b)}(n) = 0$$

2. Wenn n zu k teilerfremd ist und das Indexsystem

$$\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha, \beta,$$

besitzt, so ist

$$\chi_{(a_1,a_2,\ldots,a_r,a,b)}(n) = \rho_1^{a_1\alpha_1}\rho_2^{a_2\alpha_2}\ldots\rho_r^{a_r\alpha_r}\rho_{r+1}^{a\alpha}\rho_{r+2}^{b\beta}$$

[44] Diese h zahlentheoretischen Funktionen will ich nunmehr kurz mit

1. $\chi_1(n), \chi_2(n), \ldots, \chi_h(n)$,

bezeichnen, die allgemeine mit

$$\chi_x(n), (x = 1, \ldots, h)$$

und, wo kein Mißverständnis enstehen kann, sogar noch kürzer mit

$$\chi(n).$$

I would like to denote these h number-theoretic functions by the abbreviations

1. $\chi_1(n), \chi_2(n), \ldots, \chi_h(n)$,

the general one by the abbreviation

$$\chi_x(n), (x = 1, \ldots, h)$$

and, where no misunderstanding can arise, even shorter by

$$\chi(n)\ldots$$

The abbreviated notation, in the general case, is thus the same as Hadamard's, since here Landau indexed the characters by a natural number that provides no information about their construction data.

However, after introducing the abbreviations for the characters, Landau made the following remark [Lan09, 404] [45]:

> ...I will prove four theorems about them [the characters] with very short and elegant wording. Then the reader may soon completely forget the rather complicated definition of these functions and needs only remember that the existence of a system of h different functions which possesses the four properties has been proved.

The four theorems that Landau was referring to are the following[46] [Lan09, 401-408]:

[45] ...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, das die Existenz eines Systems von h verschiedenen Funktionen bewiesen worden ist, welche die vier Eigenschaften besitzen.

[46] **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 $n \equiv n' \pmod k$

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

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

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

Theorem 1: For any two positive numbers n, n'

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

For each of the h functions this multiplication-law holds ...
Theorem 2: For $n \equiv n' \pmod{k}$,

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

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

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

Theorem 4 When n is fixed and the sum

$$(6) \quad \sum_{x=1}^{h} \chi_x(n)$$

extends over all h functions, then

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

Satz 4: Wenn n festgehalten und die Summe

$$(6) \quad \sum_{x=1}^{h} \chi_x(n)$$

über alle h Funktionen erstreckt wird, so ist

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

dagegen

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

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

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

therefore for all $k-1$ other residue classes modulo k.

Thus Landau felt compelled to construct the characters in the same way as Dirichlet, Dedekind, Hadamard and de la Vallée Poussin, but also recognized that the four properties listed above were what was crucially important. The construction, it seems, was important for Landau only to allow us to obtain these properties.

While Landau was thus articulating a much more modern perspective in this regard, note that he was still unwilling to allow that characters to appear in the range of a summation index. For, as we see in Theorem 4 above, he adopted the same procedure as Hadamard when dealing with such sums: he permitted the summation index to range over the natural number that indexes the characters. However, unlike Hadamard, he adopted an extensional classification of the characters. He divided them as follows [47] [Lan09, 412]:

1. The first class is formed from the principal character χ_1 alone, that is always 0 or 1.
2. The second class is formed from the characters which are real for each n (therefore =0, ±1) but not always =0 or +1. In other words: they are the real characters (i.e. are always real valued) that are different to the principal character ... [Landau proceeds to give a description in terms of the construction data of the characters]

47

1. Die erste Klasse wird vom "Hauptcharakter" χ_1 allein gebildet, der stets 0 oder 1 ist.
2. Die zweite Klasse wird von denjenigen Charakteren gebildet, welche für jedes n reell (also = 0, ±1) sind, aber nicht stets =0 oder +1. Mit anderen Worten: es sind die vom Hauptcharakter verschiedenen reelen (d.h. durchweg reellen) Charakter ...
3. Die dritte Klasse wird von den "komplexen" Charakteren gebildet; darunter werden diejenigen verstanden, welche nicht durchweg reell sind, d.h. für mindestens ein n (also für mindestens eine Restklasse) nicht reell sind ... Wird für einen komplexen Charakter

$$\chi_x(n) = U_x(n) + iV_x(n),$$

wo $U_x(n)$ der reelle Bestandteil ist, so ist offenbar für ein gewissen anderen x'

$$\chi_{x'}(n) = U_{x'}(n) - iV_{x'}(n),$$

d.h. der Charakter $\chi_{x'}$ durchweg konjugiert zu χ_x.

3. The third class is formed from the “complex” characters; including those to be understood, which are not always real-valued, i.e. for at least one n (therefore for at least one residue class) they do not take a real value … Setting for a complex character

$$\chi_x(n) = U_x(n) + iV_x(n),$$

where $U_x(n)$ is the real component, then it is evident for some other x'

$$\chi_{x'}(n) = U_{x'}(n) - iV_{x'}(n),$$

i.e. the character $\chi_{x'}$ is everywhere conjugate to χ_x.

Turning now to L-functions, we see that Landau’s *Primzahlen* contains more than one perspective. Within the context of the proof of Dirichlet’s theorem, Landau introduced the L-functions as a function of a *real* variable, and again his notation is like Hadamard’s. He wrote[48] [Lan09, 414]:

> Now, for real $s > 1$, corresponding to the h characters, set the h infinite series
>
> 1. $L_x(s) = \sum_{n=1}^{\infty} \frac{\chi_x(n)}{n^s}$.
>
> Because
>
> $$\mid \frac{\chi_x(n)}{n^s} \mid < \frac{1}{n^s}$$
>
> for $s > 1$ each of these series is convergent.

In proving the non-vanishing of the L-series, he abbreviated this simply to L, when the context was clear. And, when it was necessary to sum over the different L-functions, Landau’s method was again the same as Hadamard’s: he allowed the sum to range over the index associated with each L-function. For example [Lan09, 420]:

$$\sum_{x=1}^{h} \chi_x(b) \frac{L_x'(s)}{L_x(s)} = -\sum_{p,m} \frac{\log p}{p^{ms}} \sum_{x=1}^{h} \chi_x(bp^m),$$

where b is such that $bl \equiv l \pmod{k}$.

[48] Ich betrachte nun für reelle $s > 1$, den h Charakteren entsprechend, die h unendlichen Reihen

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

Wegen

$$\mid \frac{\chi_x(n)}{n^s} \mid < \frac{1}{n^s}$$

ist jede dieser Reihen für $s > 1$ konvergent

Later in his textbook, however, Landau introduced L-functions which can take a *complex* argument, using the same notation[49] [Lan09, 458]:

> Let $s = \sigma + ti$ be a complex variable. Then the series
>
> $$L_x(s) = \sum_{n=1}^{\infty} \frac{\chi_x(n)}{n^s}$$
>
> as a Dirichlet series is convergent for $\sigma > 1$ in the case $x = 1$, and for $\sigma > 0$ in the case $x = 2, \ldots, h$ and defines there an analytic function which may be differentiated term by term.

However, despite using the notation $L_x(s)$ consistently throughout his presentation of the proof of Dirichlet's theorem and other parts of his text, Landau later switched to a more modern notation. His reason for the switch was that it was "more convenient" [Lan09, 482]. He remarked[50] [Lan09, 482]:

> Now let
>
> $$L_x(s) = \sum_{n=1}^{\infty} \frac{\chi_x(n)}{n^s}$$
>
> be that function corresponding to the character $\chi(n) = \chi_x(n)$; it is now more convenient to write the character included in the notation
>
> $$L(s, \chi),$$
>
> and when there is no fear of misunderstanding, just as before
>
> $$L(s).$$

[49]Es sei $s = \sigma + ti$ eine komplexe Variable. Dann ist die Reihe

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

als Dirichletsche Reihe im Fall $x = 1$ für $\sigma > 1$, in den Fällen $x = 2, \ldots, h$ für $\sigma > 0$ konvergent, definiert dort eine analytische Funktion und darf dort gliedweise differentiiert werden.

[50]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 krz

$$L(s).$$

Landau did not specify precisely why he found it more convenient to write $L(s,\chi)$ than $L_x(s)$. However, in the paragraph following the introduction of the new notation, Landau showed that the theory of L-functions could be reduced to L-functions that correspond to *proper characters*. Recall that, roughly, proper characters modulo k are those which cannot be obtained as a character modulo K where $K < k$. Then, to demonstrate the theory of L-functions could be reduced in the appropriate way, he showed that for an improper character (a character that is not proper) modulo k, for $\sigma > 1$, we have the following ([Lan09, 482-483]):

$$L(s,\chi) = \prod_{\nu=1}^{c} \left(1 - \frac{\epsilon_\nu}{p_\nu^s} L_0(s,X)\right),$$

where $L_0(s,X)$ is an L-function corresponding to the proper character X modulo K, ϵ_ν are roots of unity, and c a number (which can be 0, depending on how many prime factors of k are contained in K).

But let us ask whether this could be written perspicuously if Landau had not introduced his new notation. Then 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? We would have to assign a subscript to the proper character X, and thus in addition to keeping track of subscripts used to index the characters modulo k, we would need to keep track of subscripts indexing proper characters corresponding to improper characters modulo k. And given that the subscripts are just numbers, it is not clear how this could be done in a clear way. We could perhaps try to give the index itself a subscript or superscript, but this will result in messier equations that are not as efficient at presenting the salient information. And we should note that perhaps the most obvious choice for an index, x', would not be available, since this had been used previously in connection to the conjugates of characters, as noted above. These issues are compounded when, in later sections, Landau wanted to obtain functional equations to relate $L(s,\chi)$ and $L(1-s,\overline{\chi})$ and in doing so referred to the distinction between proper and improper characters (see e.g. [Lan09, §130]). Thus, in this context, using the old notation, we would need to keep track of 3 different subscripts to index the various characters, which would make the resulting work much more complicated to follow. Consequently, we may speculate that issues such as the above contributed to Landau's decision to alter his notation for the L-functions.

5.5.1 Analysis of Landau's 1909 presentation of characters and L-functions

As we have seen, Landau's *Primzahlen* contains more than one perspective concerning characters and L-functions, starting with a treatment which is very similar to Hadamard's, at least in some respects, before moving towards a much more modern approach.

Regarding the unification of the function concept, we should note that although Landau did not appear to treat the *L*-functions explicitly as functions of characters in his presentation of Dirichlet's theorem, in the later chapters, he did treat them this way. For as we have seen, he later came to allow the characters to appear in the argument position of the *L*-functions. Thus, like de la Vallée Poussin, Landau was working with a wider conception of function than Dirichlet, Dedekind or Hadamard.

With respect to the reification of the function concept, we emphasize the following points:

1. Landau gave the characters their own (constructive) definition.

2. Landau studied the characters in their own right, establishing that they enjoy four key properties, and subsequently moved towards defining them in terms of these properties.

3. The characters were given an *extensional* classification into three distinct classes.

4. Within the context of Dirichlet's theorem on primes in arithmetic progressions, the characters themselves were not permitted to appear in the range of a bound variable, but the numbers used to index them were.

5. In his presentation of Dirichlet's theorem, Landau's notation for the *L*-functions included the index of the corresponding character, and he allowed the index of the *L*-function $L_x(s)$ to appear in the range of a bound variable. This suggests Landau recognized a dependence upon the indices and, derivatively, the characters. Moreover, in the later parts of his textbook, as we have noted, Landau permitted the character to appear in the argument position of the *L*-function, thus indicating he viewed them as a function of functions.

Thus, we see that answers to all but one of our clarificatory questions ("is it permissible for the characters to appear in the range of bound variables?") are positive, and hence Landau was very close to treating the characters as legitimate mathematical *objects*.

Finally, let us remark that features of Landau's presentation of the characters and *L*-functions aid the reader in understanding the proof. We first remark that his emphasis on the properties that the characters satisfied, rather than their construction data, helps the reader identify and keep in mind the key properties required for the proof, while preventing them from becoming lost in the complicated (and mainly irrelevant) construction data. Secondly, we note that he appeared to recognize and avoid the disadvantages of adopting an approach similar to Hadamard's that can arise when working with characters and L-functions in more general settings. Indeed, he opted to change his notation for L-functions from $L_x(s)$ to $L(s,\chi)$ and, as noted above, it seems that part of his reason was doing so was to avoid obscure and confusing notation when obtaining equations relating $L(s,\chi)$ and $L(1-s,\overline{\chi})$.

Now that we have examined Landau's 1909 work, we turn to his 1927 work.

5.5.2 Landau's *Vorlesungen*

In his *Vorlesungen* of 1927, Landau moved towards a completely modern approach to characters and L-functions. Thus, in this section, we will present a very brief description of his treatment of them, focusing primarily on features which were different to his 1909 account. We adhere to the same notational conventions for the first term and common difference of an arithmetic progression, as well as for an arbitrary prime, as we did in the previous section.

Recall that, in his 1909 presentation, Landau defined the characters via the usual construction, before deriving some important properties and remarking that the reader could then forget the construction all together. In his 1927 work, however, he opted to cut out the constructive definition completely, and define the characters in terms of certain (other) properties. His introduction was as follows [Lan99, 109]:

> A number theoretic-function $\chi(a)$ is called a character $\pmod k$, provided that
> I) $\chi(a) = 0$ for $(a,k) > 1$,
> II) $\chi(1) \neq 0$,
> III) $\chi(a_1a_2) = \chi(a_1)\chi(a_2)$ for $(a_1,k) = 1$ and $(a_2,k) = 1$ (and therefore, by I), always),
> IV) $\chi(a_1) = \chi(a_2)$ for $a_1 \equiv a_2 \pmod k$ and $(a_1,k) = 1$ (and therefore, by I), whenever $a_1 \equiv a_2 \pmod k$).

And the principal character was given an entirely extensional definition [Lan99, 110]:

$$\chi(a) = \begin{cases} 0 & \text{for } (a,k) > 1 \\ 1 & \text{for} (a,k) = 1 \end{cases}$$

Thus Landau expunged all reference to the roots of unity and primitive roots modulo k from the definition of the characters, although he did build up a picture of how they were composed via different theorems, in a similar manner to how this is done in the modern proof. More specifically, he first proved that the values of the characters are roots of unity with an absolute value of 1 (see [Lan99, 109]), and later obtained the construction of the characters in his proof that for d a natural number coprime to k, there is a character modulo k, χ, such that $\chi(d) \neq 1$ (see [Lan99, 111-112]).

Correspondingly, Landau's classification of the characters was also purely extensional [Lan99, 114]:

> $\chi(a)$ is called a character of the first kind if it is the principal character; of the second kind, if it is real but is not the principal character (so that its value is always 0, 1 or -1, and -1 actually occurs); and of the third kind, if it is not everywhere real.

And, while he was not comfortable having summation indices ranging over the characters in his 1909 work, he was in his *Vorlesungen*. Indeed, he wrote the second orthogonality relation as follows [Lan99, 114]:

> For fixed $a > 0$,
>
> $$\sum_{\chi} \chi(a) = \begin{cases} c & \text{for } a \equiv 1 \pmod{k} \\ 0 & \text{for} a \not\equiv 1 \pmod{k} \end{cases}$$
>
> where the sum is taken over all of the c characters.

As to the L-functions, in the *Vorlesungen*, Landau consistently used the new notation he introduced in the later parts of his 1909 work, i.e. $L(s,\chi)$. Thus he was comfortable treating them as functions of functions. However, as in his earlier presentation of Dirichlet's theorem, these were assumed to be functions of *real* variables.

Additionally, Landau allowed the characters in the argument position of the L-functions to appear in the range of a summation index. For example, letting $h = \phi(k)$ and

$$\Lambda(a) = \begin{cases} \log p & \text{for } a \equiv p^c \; c \geq 1 \\ 0 & \text{for all other } a > 0 \end{cases},$$

he demonstrated that [Lan99, 125]:

$$-\frac{1}{h} \sum_{\chi} \frac{1}{\chi(l)} \frac{L'(s,\chi)}{L(s,\chi)} = \sum_{a \equiv l} \frac{\Lambda(a)}{a^s}$$

And, as we have argued previously this emphasizes that the L-functions are to be considered functions of the characters.

5.5.3 Analysis of Landau's *Vorlesungen*

Landau's treatment of characters and L-functions is thus practically the same as the modern approach that we examined in section 3. As such, we see that he was, as in the later parts of his 1909 text, working with a more general conception of function than some of the previous mathematicians whose work we have analyzed. Moreover, we see that there is substantial evidence to suggest that he treated the characters as fully fledged objects. More specifically, regarding the reification of the function concept, we emphasize the following points:

1. Landau gave the characters their own definition, and unlike Dirichlet, Dedekind, Hadamard or de la Vallée Poussin, he defined them in terms of their *properties*.

2. The characters were subjected to their own study: for example, Landau proved that their values were roots of unity and that they satisfied the orthogonality relations.

3. The characters were presented as independent of any representation: they were defined in terms of their properties and they were classified *extensionally*.

4. Landau allowed the characters to appear within the range of bound variables, specifically within the range of summation indices.

5. The L-functions were presented as depending upon the characters, and, moreover, being functions of the characters, since Landau utilized functional notation and allowed the functional argument to be bound by summation indices.

And, as we can see, this treatment of the characters as objects has an impact upon the ease with which the reader can understand and follow his presentation. Indeed, in contrast with his 1909 account, Landau's decision to work directly with the characters themselves instead of numbers which index them means that his equations were more perspicuous: we no longer have an intermediary, the index, that must be kept in mind when considering sums and products of characters. And the fact that the characters are defined in terms of their properties helps the reader to focus on their crucial features before coming to see how they can be constructed. Thus we do not have to attempt to work through their construction first and then "forget" this once we see that the functions satisfy certain conditions, as in his 1909 text, which appears to be somewhat of a waste of our cognitive efforts. Instead, by defining the characters in terms of the properties we wish them to have, we are not only able to follow the general proofs more easily, we can also see more clearly why the usual construction works and where it comes from. Indeed, by starting with the properties we want the characters to possess, the construction appears more natural. Moreover, as we have seen, utilizing the functional notation $L(s,\chi)$ provides us with a uniform notation for L-functions that can be adapted to more general situations, which cannot be said for Landau's previous choice of $L_x(s)$.

6 Conclusion

We have examined above the changing approaches to the characters and L-functions, identifying key features of the various treatments and the significance that they have. In this section, we conclude by making some general observations about the progression from Dirichlet's original presentation to that of Landau's and other modern approaches, as well as drawing attention to conflicting issues which may have influenced the changes.

First, let us divide the presentations of Dirichlet's theorem that we have examined into three classes. In the first class, we put Dirichlet's original treatment, in which the characters were considered only as a part of the L-functions, and nothing in their own right. In the second class, we place Dedekind, de la Vallée Poussin, and Hadamard, as well as Landau for his 1909 approach. These mathematicians, as we have seen, were willing to operate with the characters and L-functions in various ways that Dirichlet was not, but nonetheless, they were not comfortable with treating them on a par with other mathematical objects, e.g. the natural numbers. More specifically, Dedekind slipped back into referring to the construction data of the characters in places, Hadamard and Landau utilized an index notation to avoid the characters themselves appearing in the range of a bound variable, and de la Vallée Poussin refused to utilize the same summation symbol to indicate a sum ranging over the characters as one ranging over the natural numbers. In the third category, we place Landau's 1927 presentation, along with other modern approaches, such as the one we examined in chapter 3. Under these approaches, the characters were defined in terms of the various properties they satisfied and were treated similarly to e.g. the natural numbers. Indeed, Landau's 1927 definition was axiomatic, and the modern approach we examined in chapter 3 defined the characters as an extension of a special type of function (group homomorphism) that was itself defined in terms of a characteristic property .

What can we say about the transitions between these different stages? What pushed some of the mathematicians to treat the characters as mathematical objects in the sense we have explicated above, and what gave the others pause? As we have seen in the above chapters, one major issue was the intelligibility of the proof. More specifically, we identified a number of ways in which the intelligibility of the proof could be impacted in a negative way:

1. Burdening the reader with excess information. In particular, providing information about the construction of the characters that is not directly needed or used in the proof.

2. Requiring the reader to keep track of a large number of bookkeeping devices. For example, to keep in mind that a certain subscript x indexes the characters, another subscript x' indexes the conjugates of the characters and so on.

3. Requiring the reader to keep track of a large number of symbols to represent operations on different types of mathematical "things". For example, summation over numbers being denoted as $\sum_n$ but summation over characters being denoted as S_χ.

4. Lack of organization and compartmentalization of the proof. For example, not identifying the key results about the characters and proving them prior to embarking upon the proof of the main theorem.

Each of the issues above requires additional, and unnecessary, cognitive effort from the reader. Presenting the reader with more information than is required means that ultimately he must attempt to sift through it all and isolate the salient information himself. Requiring the reader to keep track of multiple book-keeping devices or symbols to represent the same operations on different "kinds" of mathematical "things" both place additional strain on the reader's memory. Finally, as discussed in Section 4.5.4, a proof that is not appropriately organized or compartmentalized will require the reader to spend additional effort in switching from following the argument of the main proof to following the proofs of various side lemmas and back again. Furthermore, a poorly organized proof will require the reader to spend additional effort to ascertain the structure of the proof, how the lemmata are utilized and why they are needed.

The negative impact of the points listed above on the readability and intelligibility of the proof is perhaps by itself sufficient reason to prompt the mathematicians whose work we have examined to adapt their presentations and gradually come to treat the characters as objects. For indeed, treating them as fully fledged objects, on a par with the natural numbers, avoids the pitfalls listed above, as we have seen by looking at presentations within the third category. However, the issues raised above come to a head when mathematicians want to utilize the characters and L-functions in more general contexts. As we have discussed in Sections 5.3 and 5.4, items 2 and 3 above become particularly pressing, since new, more general contexts require finer distinctions and thus more book-keeping devices or a further stratification of symbols used to represent certain operations. Thus retaining approaches from the first or second class renders the mathematics virtually intractable.

Given the impact that the above issues can have upon the proof, we may wonder why it appears to have taken mathematicians so long to arrive at a presentation of Dirichlet's theorem that falls into the third category. For, as we have seen, the presentations by Dedekind, de la Vallée Poussin, Hadamard and the earlier Landau all fall into the second category and as such do not treat the characters as fully fledged objects, on a par with the natural numbers, and are negatively impacted by the issues raised above.

But, we should not forget that in treating the characters as objects, and defining them in terms of their properties and not via an explicit construction, further questions arise that must be answered before the mathematics can proceed "safely". For if we have an explicit construction of the characters which we rely upon in proofs, there can be no doubt that they exist, and we can easily count them and distinguish between them based upon their representations. But, if we have a definition that involves only their properties, how do we know that there are such functions? If we have proved that there are indeed characters modulo k, how can we calculate how many of them there are? And on what grounds can we distinguish between them, if we do not know know how they are represented?

These questions are not as trivial as they may appear at first glance. Indeed, if they were not given an answer, mathematicians would have no guarantee that they had not been seduced into making mistakes in their work. As an example, recall that in the proof of Dirichlet's theorem, we must sum over all of the characters of a given modulus. Yet in order to do this, we must make sure that the sum is well defined. And if there happened to be infinitely many characters, the sum would not be well defined if it diverged, for example. However, these questions, while difficult, can be given a satisfactory answer. The modern presentation that we examined in the Chapter 3 established that group characters exist and that there are finitely many of them by appealing to Lemma (3.9) and provided a means of distinguishing between them by treating them purely extensionally. Thus all that we had to keep in mind through this presentation were their properties and their values: the explicit construction appeared in Lemma (3.9) and was referred to in establishing its corollaries, but after that the explicit construction was not at all needed or used.

Thus we see that the mathematicians working in number theory had to balance the desire to ensure that their mathematics was secure and not prone to erroneous reasoning, which could be accomplished by working with explicit constructions, with the desire to boost the readability and intelligibility of their proofs, which could be achieved by treating the characters as objects. Dirichlet seemed less concerned with the second of these two desiderata, at least for the purposes of his 1837 paper. Dedekind, de la Vallée Poussin, Hadamard and the earlier Landau, on the other hand, attempted to improve his presentation, which resulted in them taking steps towards treating the characters as objects. However, they still shied away from giving them the same status as the paradigmatic example of mathematical objects: the natural numbers. With the later Landau and other modern mathematical presentations, however, we see that the mathematicians had found a way to ensure both the security of their mathematics as well as improving its intelligibility.

References

[CFR11] C. Carstensen, B. Fine, and G. Rosenberger. *Abstract Algebra: Applications to Galois Theory, Algebraic Geometry and Cryptography*. De Gruyter, 2011.

[Cop04] W.A. Coppel. *Number Theory: An introduction to mathematics Part B*. Phalanger Press, 2004.

[DD99] P.G.L. Dirichlet and R. Dedekind. *Lectures on Number Theory*. American Mathematical Society, London Mathematical Society, 1999. English translation by John Stillwell. Originally published in German in four editions in 1863, 1871, 1879, and 1894.

[Ded88] R Dedekind. *Was sind und was sollen die Zahlen?* F.Vieweg, 1888.

[Dir29] J.P.G.L. Dirichlet. 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, 1829. Reprinted in [Kro89] pg 117-132.

[Dir37a] J.P.G.L. Dirichlet. Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält. *Bericht Ak. Wiss. Berlin*, (48), 1837. Reprinted in [Kro89], pg 315-342. English translation available as [DS08].

[Dir37b] J.P.G.L. Dirichlet. Beweis eines Satzes Über die arithmetische Progression. *Bericht über die Verhandlungen der Königl. Preuss. Akademie der Wissenschaften*, 4, 1837. Reprinted in [Kro89] pg 307-312.

[Dir37c] J.P.G.L. Dirichlet. Über die Darstellung ganz willkürlicher Functionen durch sinus und cosinusreihen. *Reportorium der Physik*, 1, 1837. Reprinted in [Kro89], pg 133-160.

[dlVP96] C.J. de la Vallée Poussin. Démonstration Simplifiée du Théorèm 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, 1896-1896.

[dlVP97] C.J. de la Vallée Poussin. *Recherches analytiques sur la théorie des Nombres Premiers*. Hayez, 1897.

[DS08] J.P.G.L. Dirichlet and R. Stephan. There are infinitely many prime numbers in all arithmetic progressions with first term and difference coprime. http://arxiv.org/abs/0808.1408, 2008.

[dWzB50] Königlich Preussische Akademie der Wissenschaften zu Berlin. *Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königl. Preuss. Akademie der Wissenschaften zu Berlin*. Königl. Akademie der Wissenschaften, 1850.

[Eul44] L. Euler. Variae observationes circa series infinitas. *Commentarii academiae scientiarum Petropolitanae*, 9, 1744.

[Eul48] L. Euler. *Introductio in analysin infinitorum, tomus primus*. Lausannae, 1748.

[EW05] G. Everest and T. Ward. *An Introduction to Number Theory*. Springer, 2005.

[Had96] J. Hadamard. Sur la Distribution des Zéros de la Fonction $\xi(s)$ et ses Conséquences Arithmétiques. *Bulletin de la Société mathématique de France*, 24, 1896.

[IK04] H. Iwaniec and E. Kowalski. *Analytic Number Theory*. Colloquium Publications - American Mathematical Society, 2004.

[Kle89] I Kleiner. Evolution of the Function Concept: a Brief Survey. *The College Mathematical Journal*, 20(4), 1989.

[Kro89] L. Kronecker, editor. *G. Lejeunue Dirichlet's Werke*. Georg Reimer, 1889.

[Lan09] E. Landau. *Handbuch der lehre von der verteilung der primzahlen*. B. G. Teubner, 1909.

[Lan27] E. Landau. *Vorlesungen über Zahlentheorie: aus der elementaren Zahlentheorie, Volume 1*. S. Hirzel, 1927.

[Lan99] E. Landau. *Elementary number theory*. American Mathematical Society, 1999.

[Luz98a] N. Luzin. Functions: Part I. *The American Mathematical Monthly*, 105(1), 1998.

[Luz98b] N. Luzin. Functions: Part II. *The American Mathematical Monthly*, 105(3), 1998.

[Mon72] A.F. Monna. 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, 1972.

[Qui61] W.V.O. Quine. *From a Logical Point of View: 9 Logico-Philosophical Essays*. Havard University Press, 1961.

[Rie67] B. Riemann. Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe. *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 13, 1867.

[Rog74] K. Rogers. Legendre's Theorem and Quadratic Reciprocity. *Journal of Number Theory*, 6(5):339 – 344, 1974.

[Sny02] N. Snyder. Artin l-functions: A historical approach. http://www.math.columbia.edu/ nsnyder/thesismain.pdf, 2002.

[Wag03] S.S. Wagstaff. *Cryptanalysis of Number Theoretic Ciphers*. CRC Press, 2003.

[You76] A. P. Youschkevitch. The Concept of Function up to the Middle of the 19 Century. *Archive for History of Exact Sciences*, 16(1), 1976.