Are there still unsolved problems about the numbers $1, 2, 3, 4, \ldots$?

Barry Mazur

(with figures by William Stein)

1 Foreword

These are notes to a talk I gave at MIT, sponsored by the Clay Mathematics Institute [see slide 2]. The talk was primarily structured around a series of over 80 mathematical slides composed by William Stein (along with some others made by Bill Casselman). William Stein and I have the intention of publishing a pamphlet, or short book, containing a slight elaboration of this text with figures, accompanied by a CD containing all the slides. The current text, then, is a very preliminary draft of our forthcoming pamphlet. Its mission is to explain what is exciting about the Riemann Hypothesis, using no formulas and very very little mathematical vocabulary: our main tool will be graphs. For this to work, then, a reader should watch the accompanying slides, where it is indicated to do so in the text.

Here are the mathematical prerequisites. We will want our readers to be able to understand graphs, such as the bar graphs that occur in newspapers. The two letters we reserve to indicate numbers are N (which can stand for any positive whole number, $N = 1, 2, 3, \ldots$) and X (which can stand for any positive real number, whole, or fractional, or general). The *main* graph we will focus on is something that we denote $P(N)$: the height of the graph $P(N)$ above any positive whole number N is the number of prime numbers less than or equal to N. All this will be discussed more precisely in the text below. We call this graph *the staircase of primes*[1]. We have labeled it $P(N)$, since this graph is entirely determined if you know its height above the positive whole numbers N. Most of the rest of our graphs will be smooth curves, and these we label $G(X)$, $R(X)$, etc. because to specify such a graph you must give its height above any real number X, and not just above whole numbers.

Beginning in section 20, it would useful to know that *complex numbers* are of the form $x + iy$ where x and y are real numbers; the real number x is usually referred to as the **real part** of the complex number $x + iy$, and y is called its **imaginary part**. A complex number $x + iy$ is usually represented as the point (x, y) in the Euclidean plane.

There are four brief endnotes and these need *not* be consulted to understand the text; they are only meant as a bit of a convenience for readers who know calculus, and who might wish to know some more technical aspects of the subject under discussion.

[1] borrowing a phrase from Du Sautoi's book on the Riemann Hypothesis

2 Thoughts about numbers: ancient, medieval, and modern

If we are to believe Aristotle, the early Pythagoreans thought that the principles governing Number are "the principles of all things," the elements of number being more basic than the Empedoclean physical elements *earth, air, fire, water*. To think about number is to get close to the architecture of " what is." So, how far along are we in our thoughts about numbers?

René Descartes, almost four centuries ago, expressed the hope that there soon would be "almost nothing more to discover in geometry." Contemporary physicists dream of a "final theory." But despite its venerability and its great power and beauty, the pure mathematics of numbers may still be in the infancy of its development, with depths to be explored as endless as the human soul, and never a final theory.

Numbers are obstreperous things. Don Quixote encountered this when he requested that the "bachelor" compose a poem to his beloved Dulcinea del Toboso, the first letters of each line spelling out her name. "To be her slave and hers alone, nature threw me into this world," Don Quixote sighed, despite the fact that any reader of Cervantes has real reason to wonder whether Don Quixote had actually ever seen the lady. The "bachelor" found it difficult to compose such a poem

> because the number of letters in her name was 17, and if he made four Castilian stanzas of four octosyllabic lines each, there would be one letter too many, and if he made the stanzas of five octosyllabic lines each, the ones called *d*écimas or *r*edondillas, there would be three letters too few...

Not willing to grant the imperviousness of the number 17 to division, Quixote pleaded:

> It must fit in, however, you do it.

Seventeen is indeed a prime number: there is no way of factoring it as the product of smaller numbers, and this accounts–people tell me– for its occurrence in some phenomena of nature, as when last year the 17-year cicadas all emerged to celebrate a "reunion" of some sort in our fields and valleys.

Prime numbers are mentioned as a class of numbers in the writings of Philolaus (a predecessor of Plato); they are not mentioned specifically in the Platonic dialogues, which is surprising to me given the intense interest Plato had in mathematical developments; and they make an occasional appearance in the writings of Aristotle, which is not surprising, given Aristotle's emphasis on the distinction between the *composite* and the *incomposite*. "The incomposite is prior to the composite," writes Aristotle in Book 13 of the Metaphysics (thereby making–among other things–quite an interesting atemporal use of the word "prior").

Prime numbers, despite their *primary* position in our modern understanding of number, were not specifically doted over in the ancient literature before Euclid, at least not in the literature that has been preserved. Until Euclid, prime numbers seem not to have been singled out as *the* extraordinary mathematical concept, central to any deep understanding of numerical phenomena, that they are now understood to be.

It is easy to answer the rhetorical question of the title of this lecture: there is an extraordinary wealth of established truths about numbers; these truths provoke sheer awe for the beautiful complexity of prime numbers. But each of the important new discoveries we make give rise to a further richness of questions, educated guesses, heuristics, expectations, and unsolved problems.

Central to the mystery of prime numbers lies one of the great unsolved problems of mathematics, The Riemann Hypothesis, a problem that offers $1,000,000 of *Clay Mathematics Institute* prize

money for the person who solves it, and–with or without money– its answer is crucial for our understanding of the nature of numbers. I will try to explain what the Riemann Hypothesis is, and why its resolution will give us such important insight into the "deep structure" of primes. As encouragement for this journey I can do no better than to quote a paragraph of Don Zagier's classic 12-page exposition of all this mathematics, *The First 50 Million Prime Numbers*:

> There are two facts about the distribution of prime numbers of which I hope to convince you so overwhelmingly that they will be permanently engraved in your hearts. The first is that, [they are] the most arbitrary and ornery objects studied by mathematicians: they grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout. The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behavior, and that they obey these laws with almost military precision.

3 Primes as atoms.

To begin from the beginning, think of the operation of multiplication as a bond that ties numbers together: the equation $2 \times 3 = 6$ invites us to imagine the number 6 as (a molecule, if you wish) built out of its smaller constituents 2 and 3. Reversing the procedure, if we start with a whole number, say 6 again, we may try to factor it (that is, express it as a product of smaller whole numbers) and, of course, we would eventually, if not immediately, come up with $6 = 2 \times 3$ and discover that 2 and 3 factor no further; the numbers 2 and 3, then, are the indecomposable entities (atoms, if you wish) that comprise our number. By definition, a **prime number** (colloquially, *a prime*) is a whole number, bigger than 1, that cannot be factored into a product of two smaller whole numbers. So, 2 and 3 are the first two prime numbers. The next number along the line, 4, is not prime, for $4 = 2 \times 2$; the number after that, 5, is. Primes are, multiplicatively speaking, the building blocks from which all numbers can be made. A fundamental theorem of arithmetic tells us that any number (bigger than 1) can be factored as a product of primes [slide 4] and the factorization is *unique* except for rearranging the order of the primes. For example, if you try to factor the number 300, there are many ways to begin:

$$300 = 30 \times 10$$

or

$$300 = 6 \times 50$$

and there are various other starting possibilities. But if you continue the factorization ("climbing down" any one of the possible "factoring trees") to the bottom, where every factor is a prime number, e.g.,

$$300$$

$$30 \times 10$$

$$3 \times 10 \times 2 \times 5$$

$$3 \times 2 \times 5 \times 2 \times 5$$

you always end up with the same collection of prime numbers:

$$300 = 2^2 \times 3 \times 5^2.$$

[slides 5-11]

The uniqueness of the ultimate collection of prime factors, the atoms, of any number N– independent of all the choices that may be open to you when you are proceeding with the tree of factorizations of N–is a genuine theorem, requiring proof. Even though this uniqueness result was used by earlier mathematicians, Carl Friedrich Gauss, at the beginning of the nineteenth century, was the first to give an explicit proof of it.

The Riemann Hypothesis will probe the question: how intimately can we know prime numbers, those *atoms* of multiplication?

4 Large primes in the real world

Anytime we visit an e-commerce website, prime numbers having a hundred digits (or more) are used to keep our on-line bank transactions private. This ubiquitous use to which these giant primes are put depends upon a very simple principle: it is much easier to multiply numbers together than to factor them. If you had to factor, say, the number 143 you might scratch your head for a few minutes before discovering that 143 is 11×13. But if you had to multiply 11 by 13 you'd do it straightaway. Offer two primes, say, P and Q each with more than 100 digits, to your computing machine, and ask the computer to multiply them together: you will get their product $N = P \times Q$ with its 200 (or so) digits in nanoseconds. But present that number N to any current computer, and ask it to factor N– and the computer will fail to do the task. The safety of much encryption depends upon this failure! (But we have no theorem, yet, guaranteeing us that factoring is necessarily as difficult as we presently think it is.) Every time you send secure e-mail, say, to your bank, the bank's computer will be doing the following litany of things, for each interaction it makes with you. The bank will

1. generate (by a process we will discuss below) a pair of large (i.e., hundred digit) prime numbers P and Q,
2. multiply them together, $N := P \times Q$, and
3. communicate *only* the (roughly) 200-digit number N to you (more exactly: to your computer) together with certain instructions about how to encode the message you wish to send to the bank; the encoding process makes use of this number N, is easy for your computer to perform, and is virtually impossible for anyone else to decode without their first factoring the number N into its prime constituents P and Q.

This scenario requires, among other things, that there is a feasible way for the bank to get a reliable and plentiful supply of 100-digit prime numbers. In fact, there is such a way, and this is already a minor miracle: As we shall be discussing in more detail later in this pamphlet, the probability that a number of that size is prime is roughly 1/230. Moreover, if, in your random process, you build in a groundrule that avoids numbers that are divisible by a few small primes, you can greatly reduce these odds. For any number P, there are tests that are rapid to administer and that are discerning enough, so that if P passes these tests, it is virtually certain that the number is prime. So the bank's computer will be choosing these candidate-primes P's by just choosing 100 digit numbers that aren't divisible by a few small primes at random, testing them, and chucking them out if they fail the primality test. Since computers can do this pretty fast, and since, by our estimates, we can expect to get a "keeper" P better than 1 out of only 230 trials, the computers can amass, within microseconds, as many of these hefty prime numbers P as is needed. (Recently, [Agrawal et al] produced a fast method that is a *definitive* test for primality. But to guarantee the privacy of your communication, the banks only need *virtual certainty* that the P and Q used in the encoding process are primes, and not *mathematical certainty.*)

In a word, there are loads of 100 digit prime numbers, and the bank can find them, and make abundant practical use of them, to keep your transactions safe.

5 The largest prime number

There are infinitely many prime numbers. Euclid's *Elements* offers this ingenious way of seeing that any finite list of primes,

$$2, 3, 5, 7, 11, \ldots, p,$$

doesn't contain all primes: Just multiply all those primes together and add 1, to get a number

$$N = 2 \times 3 \times 5 \times 7 \times 11 \times \cdots \times p + 1.$$

Now, like all numbers bigger than 1, this constructed number N is either prime itself or is divisible by *some* prime number. But N is certainly not divisible by any of the primes up to p, because by its very construction, the number N has a remainder of 1 when divided by any of those primes. So, every prime dividing N must be a *new* prime. In a word, no finite list of primes is complete!

The clarity of this argument, however, should not make us think we have here a usable recipe for concretely finding larger, and ever larger, collections primes. It starts promising enough: if you give Euclid, say, the first two primes $2, 3$, Euclid will construct for you the number $N = 2 \times 3 + 1$ which is indeed the prime 7. If you give him the first three primes $2, 3, 5$ he will offer $N = 2 \times 3 \times 5 + 1$ which is the prime 31. But once you have produced the first two hundred or so prime numbers, the gargantuan N (the product of all those primes plus 1) that Euclid will be offering you as a number whose prime factors are new, is such a large number that neither you nor your computing machine will be able to factor it, to concretely obtain those new primes.

It is not surprising, then, that some people have worked hard to find –in explicit terms– very large prime numbers. The world's record-holder to date is the number

$$P = 2^{25964951} - 1$$

that has $7,816,230$ digits! and was proved to be prime in February 2005. If you google this number, you will learn the story of the discovery of its primality, a joint venture by number-enthusiasts around the world. You will learn, for example, that to deal effectively with primality issues regarding such a number (of close to eight million digits) you better to have a customized theory, specially designed to treat it: this is a number so large that it needs a whole *theory of its own!*

You might wonder about the peculiar way that our record-holding prime number is given: $2^{25964951} - 1$. For us to be able to handle numbers this large, they had better come along with generous hints about how to compute themselves. This one does, whispering to us: multiply 2 by itself a certain number of times and then subtract 1. Of course, even our standard decimal notation offers a computational procedure for producing the denoted number: when we write, say, "389," we are recommending that we add three hundreds to eight tens to nine ones to get the number to which we are referring.

Numbers of the above form, $2^n - 1$, are called a Mersenne numbers, after the the Minimite friar Marin Mersenne, who lived in the beginning of the seventeenth century. For such a number, $2^n - 1$, to have a fighting chance of being prime, the exponent n better be a prime itself, and–indeed– $n = 25964951$ is prime.

You might also wonder what exactly it means to say that this prime number $P = 2^{25964951} - 1$ is trumpeted as "explicitly given." Let Q denote the smallest prime number larger than P, which exists, by Euclid's theorem that we recalled above, and which is a perfectly well-defined number. What are the qualitative differences between the way in which the prime number P was presented to us, and the prime number Q? Isn't Q "explicitly given" as well? It is a good idea, by the way, to keep wondering about this: lots of mathematics can be understood by pondering this issue. Here is a tiny indication about why I, at least, think that there is a serious difference between the way in which P and Q are given. I can, if you asked me, answer lots of specific questions about P that I cannot answer about Q; for example, by an easy computation I see that the last digit in the decimal expansion of P is 7, whereas I'm willing to bet that no one, in my lifetime, will ever be able to tell me what the last digit of the decimal expansion of Q is.

6 Sifting numbers to find primes

Eratosthenes, the great mathematician from Cyrene (and librarian at Alexandria, and correspondent of Archimedes) explained how to *sift* the prime numbers from the series of all numbers: in the sequence of numbers,

2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29,

for example, start by circling the 2 and crossing out all the other multiples of 2. Next, go back to the beginning of our sequence of numbers and circle the first number that is neither circled nor crossed out (that would be, of course, the 3); then cross out all the other multiples of 3. This gives the pattern: go back again to the beginning of our sequence of numbers and circle the first number that is neither circled nor crossed out; then cross out all of its other multiples. Repeat this pattern until all the numbers in our sequence are either circled, or crossed out, the circled ones being the primes.

Especially if you have had no experience with math, may I suggest that you actually follow Eratosthenes' lead, and perform the repeated circling and crossing-out to watch the primes emerge, intriguingly staggered through our sequence of numbers,

2 3 • 5 • 7 • • • 11 • 13 • • • 17 • 19 • • • 23 • • • • • 29, . . .

Once you have done this, please go through the slides 13-21; this sequence of slides gives an animated view of the sifting process, leading to a list of all prime numbers among the first 160 natural numbers. The slide 13 displays all the numbers from 1 to 160 in a rectangular array; ignore the rectangularity of the array and think of it as just a linear sequence of numbers. Slide 14 will show the number 1 as blue–marking it for elimination, since it is not a prime. Slide 15 will have eliminated the numeral 1, highlighted the next number (the numeral 2) indicating that it is a prime, and marking all the higher multiples of 2 blue for subsequent elimination. Slide 16 will have eliminated the numbers colored blue in the previous slide, highlighted the lowest surviving number (the numeral 3) indicating that it is a prime, and marking all the higher multiples of 3 blue for subsequent elimination; etc.

7 Seemingly simple questions about the spacing of primes

We become quickly stymied when we ask quite elementary questions about the spacing of the infinite series of prime numbers.

For example, *are there infinitely many pairs of primes whose difference is* 2*?* The sequence on the page seems to be rich in such pairs

$$5-3=2, \quad 7-5=2, \quad 13-11=2, \quad 19-17=2,$$

and we know loads more such pairs (for example, there are $1,177,209,242,304$ such pairs less than $1,000,000,000,000,000$) but the answer to our question, *are there infinitely many?*, is not known. *Are there infinitely many pairs of primes whose difference is* 4*?* Answer: equally unknown. *Is every even number greater than* 2 *a sum of two primes?* Answer: unknown. *Are there infinitely many primes which are* 1 *more than a perfect square?* Answer: unknown.

Is there some neat formula giving the next prime? More specifically, *If I give you a number N, say N = one million, and ask you for the first number after N that is prime, is there a method*

that answers that question without, in some form or other, running through each of the successive numbers after N rejecting the nonprimes until the first prime is encountered? Answer: unknown.

These are questions that have been asked about primes (and we could give bushels more), questions expressible in simple vocabulary, that we can't answer today. We have been studying numbers for over two millenia and yet we are indeed in the infancy of our understanding.

8 A closer look at the sieve of Eratosthenes

Slow as we are to understand primes, at the very least we can try to count them. You can see that there are 10 primes less than 30, so you might encapsulate this by saying that the chances that a number less than 30 is prime is 1 in 3. This frequency does not persist, though; here is some more data:

There are 168 primes less than a thousand (so we might say that among the numbers less than a thousand the chances that one of them is prime is roughly 1 in 6).

There are $78,498$ primes less than a million (so we might say that the chances that a random choice among the first million numbers is prime have dropped to roughly 1 in 13).

There are $455,052,512$ primes less than ten billion; i.e., $10,000,000,000$ (so we might say that the chances are down to roughly 1 in 22).

Primes, then, seem to be thinning out. We will return to this with more specific predictions later, but for now, let us accompany Eratosthenes for a few further steps in his sifting process to help us get an initial sense of this "thinning out." Slides 22-24 will give us a picture of how the sieving process works after eliminating the higher multiples of 2 (slide 22), of 2 and 3 and then 2, 3, and 5 (slide 23) and of 2,3,5, and 7 (slide 24).

The red curve in these figures actually counts the primes: it is the beguilingly irregular *staircase of primes.* Its height above any number N on the horizontal line records the number of primes less than or equal to N, the accumulation of primes up to N. Refer to this number (of primes less than or equal to N) as $P(N)$. So $P(2) = 1$, $P(3) = 2$, $P(10) = 4$, $P(30) = 10$.

Consider slide 22, which will show us how many numbers remain, after the first pass of Eratosthenes' sieve, panning for primes. The bar graph in slide 22 that looks like a regular staircase, each step the same length as each riser, climbing up at, so to speak, a 45 degree angle, simply counts all numbers up to and including N. So its height above the number 10 is 10, above 20 is 20, etc.

The first of Eratosthenes' moves was to throw out roughly half of these numbers (the even ones!) after the number 2. The bar graph labeled *sieve 2* in this figure records this; it graphs the numbers that *are left* after this first pass through Eratosthenes' sieve, i.e., after sifting out all higher multiples of 2; so it is, with one hiccup, a regular staircase climbing at a smaller angle, each step twice the lengths of each riser. These remaining numbers include, of course, all the primes. Visibly, then, the chances that a given number in any interval of large numbers is prime is *at most* 1 in 2.

Our second move was to throw out all multiples of 3 that are bigger than 3; that is, all the ones that have survived the previous pass of Eratosthenes' sieve. Every even multiple of 3 had already been eliminated, so we are throwing out–roughly– 1/6 of the numbers in this pass. The bar graph labeled sieve 3 in Slide 23 illustrates this. These remaining numbers include, again, all the primes. So, the "probability" that a given number in any interval of large numbers is prime is at most $1 - \frac{1}{2} - \frac{1}{6} = 1/3$.

And so it goes: with each Eratosthenian move in our sieving process we are (to radically change our metaphor) winnowing the field more extensively, reducing the chances that the later numbers are prime, as recorded for the next two passes of Eratosthenes' sieve in slide 24.

This data may begin to suggest to you that as you go further and further out on the number line the percentage of prime numbers among all whole numbers tends towards 0% (it does). We will return to this.

9 Counting the prime numbers less than N

To get a sense of how the primes accumulate, we will take a look at the staircase of primes for different ranges (up to $N = 100$ and 500 in slide 25; and up to $N = 1000$ and $10,000$ in slide 26).

The striking thing about these figures is that as the numbers get large enough, the jagged accumulation of primes, those quintessentially discrete entities, becomes smoother and smoother, to the eye. How strange and wonderful to watch, as our viewpoint zooms out to larger ranges of numbers, the accumulation of primes taking on such a clean, elegant, shape. But don't be fooled by the seemingly smooth shape of the bottom curve in slide 26: it is just as faithful a reproduction of the staircase of primes as the typographer's art can render, for there are actually $1,229$ tiny steps and risers in this curve, all hidden by the thickness of the print of the drawn curve in the figure.

Someone once said, I've forgotten whom[2], that "a mathematician is a person on whom nothing is lost." This being so, no mathematician can gaze at the bottom curve of slide 26 without asking for some elegant description, a formula perhaps, for "it," or at least, for some approximation of it. A Cinderella-like project, then: can we find a well-fitting smooth curve? That last question does

not necessarily have a single answer. If I draw a curve with chalk on the blackboard, this can signify a myriad of smooth (mathematical) curves all encompassed within the thickness of the chalk-line, all–if you wish–reasonable approximations of one another. So, there are many smooth curves that fit the chalk-curve. With this warning that there may be many answers forthcoming, but fortified by the data of slide 26, let us ask: *is there an easily describable smooth curve that is a reasonable approximation to the staircase of primes?*

10 Pure and applied mathematics

At this point, I hope I don't wear down your patience, for I would like to indulge in a digression about the role of question-asking (as above) in mathematics. Loosely speaking, there are two types of mathematics: *pure* and *applied*. Now, I probably put more theoretical math in the *applied* category than most of my colleagues would. This is because I want a given piece of mathematics to be categorized as "pure" or "applied" turning on the intention behind it: if it is meant to eventually be used elsewhere (either within or outside the discipline of Mathematics) we'll call it *applied*, while if it is meant to explain and get at the root of some concept, we'll call it *pure*. Of course, there is a great overlap. Moreover, many questions in mathematics are "hustlers" in the sense that, at first view, what is being requested is that some simple task be done (e.g., *to find a smooth curve that is a reasonable approximation to the staircase of primes*). Once successfully done, it will, presumably, have many useful by-products–that is, it will be a piece of mathematics that I might classify as *applied*. But, the fuller motivation behind the question is *pure*: to strike behind the mask of the appearance of the mathematical situation, and get at the hidden fundamentals that actually govern the phenomena.

[2] its not Jonathan Swift

The particular issue before us is, in my opinion, twofold, both applied, and pure: can we curve-fit the "staircase of primes" by a well approximating smooth curve? The story behind this alone is marvelous, has a cornucopia of applications, and we will be telling but the tiniest piece of it below. Our curiosity, though, is also driven by a question that is pure, and less amenable to precise formulation: are there mathematical concepts at the root of, and more basic than (and "prior to," to borrow Aristotle's use of the phrase,) *prime numbers*–concepts that account for the apparent complexity of the nature of primes?

11 The probability that a given number is prime: a first view of Gauss's guess

Carl Friedrich Gauss, in a letter written in 1849, claimed that as early as 1792 or 1793 he had already observed that the density of prime numbers over intervals of numbers of a given rough magnitude X seemed to be proportional to

$$\frac{1}{\text{the number of digits of } X}.$$

You might get a glimmer of the thinking that supported this guess by recalling how the sieve of Eratosthenes works: for a number to be prime, the larger it is, the more sifting operations it must "survive." So the probability that numbers are prime is, one would imagine, declining as the size of the number in question grows.

Gauss's guess, though, is a good deal more precise than what was written in the previous paragraph, and here is our first way of paraphrasing it[3]

As D goes to infinity, the ratio:

$$\frac{D \times \{\text{the number of all } \textit{prime} \text{ numbers having } D \text{ digits}\}}{\{\text{the number of all numbers having } D \text{ digits}\}}$$

has as limiting value the number[4]

$$1/2.30258509299\ldots \;=\; 0.434294481903\ldots$$

This tallies with the estimate we quoted earlier, that the probability of a hundred-digit number being prime is about $1/230$. But it also suggests that the probability of a thousand-digit number being prime is about $1/2302$, etc.

Gauss was an inveterate computer: he wrote in his 1849 letter that there are $216,745$ prime numbers less than three million (William Stein tells me that this is wrong: the actual number of these primes is $216,816$). Gauss's specific guess for this range predicted that there would be $216,971$ primes– a miss, Gauss thought, by 226 (but actually he was closer than he thought: the correct *miss* is a mere 161; not as close as the recent US elections, but pretty close nevertheless).

Another way of visualizing Gauss's guess is to draw the smooth curve that gives his suggested probability, for each number X, of X being prime[5] and to note that the number of primes in any reasonable range should be well approximated by the area under this curve, over that range. See slide 27 which illustrates this.

[3] Nowadays it goes under the name *The Prime Number Theorem;* yes, Prime Number "Theorem," for it was eventually proved, a century after Gauss hinted that it might be true. See further discussion about this in section 23 below.

[4] $1/\log(10)$

[5] $1/\log(X)$

12 The search for smooth curves that are "close-fits" to the staircase of primes

The full-fledged search for such approximating curves began two centuries ago with "Gauss's guess," described in the previous section. His guess leads to a certain beautiful curve that, experimentally, seems to be an exceptionally good fit for the staircase of primes. Let us denote Gauss's curve[6] $G(X)$; this curve has an elegant simple formula comprehensible to anyone who has had some of calculus [see the endnote 1] but if you want to visualize it in the range of numbers between, say, 1 and $10,000$, take *another look* at the bottom curve of slide 26. This curve is the staircase of primes, but Gauss's curve that we are denoting $G(X)$ *fits within the pen-line of the red graph* it is such a close approximation to $P(N)$.

Half a century after Gauss's initial exploration, Bernhard Riemann produced a slight variant of Gauss's approximating curve [see endnote 1] which has some heuristic reason to be an ever-so-small improvement [one can explain Riemann's curve also by consideration of probability]. Riemann's curve, which we will denote $R(X)$, seems to predict the actual number of primes to an astounding accuracy. For example, there are precisely $50,847,534$ prime numbers under a billion and Riemann's curve would predict a number that undercounts this $50,847,534$ by a tiny 79. Being an even closer fit than Gauss's, Riemann's $R(X)$ settles, with even greater ease, within the pen-line of the bottom graph of slide 26.

But we are getting ahead of ourselves, because we must discuss what we mean by "close fit."

13 How close is close?

Raoul Bott once said that whenever you read a mathematics book, or go to a math lecture, it is a good idea if you come home with something (it can be small, but should be definite) that has application to a wider class of mathematical problem than was the focus of the lecture. Since we are about to introduce criteria of *closeness of fit* and *good approximation* I can't resist suggesting that if you never have had the occasion to think about useful ways of framing the concept of *goodness* for approximating magnitudes, the criterion we are about to review seems to crop up all over mathematics, so is a fine "something" to mull over.

We will begin by formulating a rough-and-ready concept of closeness. If you are trying to estimate a number, say, around ten thousand, and you get it right to within a hundred, I propose that we celebrate this kind of accuracy by saying that you have made an approximation with *square-root error* ($\sqrt{10,000} = 100$). Of course, I should really use the more clumsy phrase "an approximation with at worst *square-root error*." If you are trying to estimate a number in the millions, and you get it right to within a thousand, let's agree that–again– you have made an approximation with *square-root error*. So, when Gauss thought his curve missed by a mere 226 in estimating the number of primes less than three million, it was well within the margin we have given for an approximation with *square-root error*.

More generally, if you are trying to estimate a number that has D digits and you get it almost right, but with an error that has no more than, roughly, half that many digits, let us say, again, that you have made an approximation with *square-root error*.

Square-root error is very sharp. It is hardly ever achieved, for example, in the social sciences, or in population statistics of large numbers. You can see this for yourself if you go to the US Census web-page. You can read, for example, that in November 2000 there were $2,957,000$ unemployed males in the civilian labor force, with a standard error of $91,000$. Setting aside the issue of how

[6] it is usually called $Li(X)$, standing for "logarithmic integral."

well-defined the phrase "civilian labor force" is, if it were accurate up to square root error, the tolerance would be a mere $2,000$ as possible error.

In the empirical sciences, approximation up to square root error is, pretty much, the gold standard: one cannot expect to do better, and it's great when this precision can be accomplished.

In contrast, in mathematics, if our expectation is that the formulas we write down are "square root close" to the data we are studying, we haven't finished our job until we actually prove this, or find ourselves forced to change our expectation. We are surely missing some essential, and perhaps structural, understanding until such an issue is resolved.

This being said, we will be specifically interested not in approximating just a *single* number, but rather all values $P(N)$ (the staircase of primes) for arbitrarily large numbers N. For this, we want to frame the concept of "square root accuracy" to have a bit of flexibility as N tends to infinity. If we are given an approximation $P'(N)$ to $P(N)$ where the error between these values is of size less than

$$\sqrt{N} = N^{0.5}$$

then we would have that our $P'(N)$ is indeed precisely "square-root close" as we have literally described this concept above. But we wish to have something slightly looser:

If for any exponent bigger than 0.5 *(e.g.,* 0.501 *or* 0.5000001*, etc.) the size of the error between our* $P'(N)$ *and* $P(N)$ *is less than* N *raised to that exponent for sufficiently large* N*, we'll happily say that our* $P'(N)$ *is "square-root close" to* $P(N)$*.*

One of the virtues of this revised notion of square root close is that it is an *equivalence relation* [see endote 2]. In particular, if you have three graphs, the first graph being square root close to the second, and the second graph square root close to the third, then all three graphs are square root close to each other.

14 What is Riemann's Hypothesis? A first view.

The elusive Riemann Hypothesis takes its origin from some awe-inspiring, difficult to interpret, lines in Bernhard Riemann's magnificent 8-page paper, "On the number of primes less than a given magnitude," published in 1859. See slide 77 for a photo of the first page of Riemann's handwritten text.

There are many seemingly different–yet equivalent– formulations Riemann's hypothesis, and our first visit to it will be in terms of the notion of closeness as introduced at the end of the previous section.

Note that square-root closeness is fundamentally an *asymptotic* notion, having to do with behavior as numbers get arbitrarily large, and since we are mathematicians, no *finite* amount of data–no matter how convincing–can definitively determine the truth or falsity of such an asymptotic statement, and so cannot answer the following questions for us:

Is Gauss's guess $G(X)$ square-root close to $P(N)$?

Is Riemann's guess $R(X)$ square-root close to $P(N)$?

It turns out to be an easy matter to check that Gauss's guess $G(X)$ and Riemann's guess $R(X)$ are square-root close to each other. Therefore to answer either one of the questions we have displayed above is to answer both of them. Moreover, the Prime Number Theorem, that we initially encountered in section 11 guarantees that these curves $G(X)$ and $R(X)$ are not wildly off the mark, for one way of stating the essential content of the Prime Number Theorem is that the ratio of $G(N)$ to $P(N)$ tends to 1 as N goes to infinity. (Or, equivalently, that the ratio of $R(N)$ to $P(N)$ tends to 1 as N goes to infinity.)

Riemann's famous hypothesis can be paraphrased as saying that the answer to these two questions is yes:

The Hypothesis of Riemann (first view): *Gauss's $G(X)$ and Riemann's $R(X)$ are both square-root close to $P(N)$, the staircase of primes!*

A more visual, but mathematically identical, way of expressing Riemann's hypothesis is that the estimate you get for the number of primes less than X by simply computing the area under the graph of slide 27 over the stretch from 2 to X is accurate to within *square-root error.*

This Riemann Hypothesis remains unproved to this day, and therefore is "only a hypothesis," as Osiander said of Copernicus's theory, but one for which we have overwhelming theoretical and numerical evidence in its support. Moreover, it turns up as relevant, as a key, again and again, in different parts of the subject: if you accept it as *hypothesis* you have an immensely powerful tool at your disposal: a mathematical magnifying glass that sharpens our focus on number theory. It is the kind of conjecture that Frans Oort and I want to label a *suffusing conjecture* in that it has unusually broad implications: many many results are now known to follow, if the conjecture, familiarly known as "RH," is true. A proof of RH would, therefore, fall into the *applied* category, given our discussion above. But however you classify RH, it is a central concern in mathematics to find its proof (or, its disproof!).

We shall see a different, and somewhat more astounding –nevertheless equivalent–way of stating Riemann's hypothesis shortly.

15 Fourier's Theory: analysis and synthesis

Gauss made his guess in the 1790s; Riemann made his in 1859, half a century later. Much mathematics had been discovered in the intervening half century. As we shall see, Riemann went much further than simply finding an approximation to the staircase of primes. To appreciate the format of Riemann's inquiry, it may pay to make a digression to discuss the 1822 contribution of J.-B. Fourier: the treatise *Théorie analytique de la Chaleur* that, apparently, Lord Kelvin referred to as a "mathematical poem." [Slide 32]

Fourier's viewpoint is now used throughout the scientific world.

When applied to *acoustics* Fourier's theory will aid in analyzing and/or synthesizing any sound: it will analyze the sound wave into its component pure frequencies. Also–if you are so disposed—you can synthesize that sound wave by recomposing it, given the information of its constituent frequencies, and their amplitudes.

If, say, the sound emitted has the shape of a "square wave" on an oscilloscope, as depicted in slide 33, J.-B. Fourier will give you the tools to synthesize this sound–or, indeed, any periodic auditory sound– by superposition of certain pure tones. Here is the basic vocabulary for this, but I will describe it specifically with reference to the square wave. The square wave is–as depicted in slide 33–periodic and has a fundamental frequency. The main approximant to this wave is a pure tone (a "sine wave") of that same frequency; this pure tone is aptly referred to simply as **the fundamental**. To get better and better approximations to a periodic sound wave, we need only superimpose the *harmonics* of the sound wave, these being pure tones whose frequencies are multiples of the fundamental frequency, each possibly shifted in time (this shift being called its **phase**) and each coming with a specific amplitude (i.e., *size*). All these harmonics, each with its frequency, phase, and amplitude, are dictated to us by an easy recipe, given in Fourier's treatise. These pure tones of higher frequencies that we need for our better approximations are called the **harmonics**

and, in referring to them, I will just rank them by increasing frequency, calling the harmonic whose frequency is closest to the fundamental frequency the **first harmonic**, the next one the **second harmonic** and so on. I suppose we might refer to the *fundamental* as the **zero-th harmonic**, but I'll resist this temptation. The collection of frequencies of the harmonics that compose the wave being synthesized is called the **spectrum** of the wave. Slide 33 shows the approximations to the square wave that one gets starting with the fundamental, and adding one, two, and four, harmonics. If we go *all the way* adding the infinitely many harmonics prescribed by Fourier's theory, we get a perfect fit. Fourier's theory works like a charm for our square wave: its first harmonic is 3 times the fundamental frequency and its *amplitude* is 1/3 the amplitude of the fundamental, its second harmonic is 5 times the fundamental frequency and has 1/5 the amplitude of the fundamental, etc. And all the harmonics are "in phase" with the fundamental. So the spectrum of our square wave consists of all odd multiples of its fundamental frequency. [You can hear a few of these approximations by playing the audio CD.]

If you listen to the audio CD you'll hear successive approximations to the square wave sound (which– at the right frequency–resembles an anemic door buzzer). You'll also hear a repeat of this sequence, to symbolize the fact, known to all engineers, that to get the various Fourier approximants to a specific periodic sound you might build it up–synthesize it–by superimposing more and more of its harmonics, if you happen to know them; or, even if you don't, you can subject the sound to a variety of bandpass filters, breaking it down, so to speak, into its constituent frequencies, and then recreating the sound, or approximations to it, from them. This dual attitude towards Fourier's ideas–analyzing a wave into the spectrum of frequencies that comprise it, or *synthesizing* the wave, i.e., building it up, by composing its *fundamental* with the various *harmonics* that comprise it– will be relevant in our discussion of Riemann's ideas.

16 Signature and Key

Fourier's theory has settled in as the very language of optics, and electromagnetic phenomena: any source of light is understood (analyzed) as comprised of its **spectrum,** i.e., the sum of pure frequencies–with their corresponding phases and amplitudes–that make up that light source. Such a spectral analysis of, say, the light emitted from the Hydrogen atom, is both

• the *signature* of the hydrogen atom, in the sense that it is so characteristic of hydrogen that form the spectrum alone you can tell if it is, or isn't hydrogen that you are dealing with, and

• a crucial *key* to the deeper internal structure of hydrogen: the spectral lines are distanced, one from another, in proportion to the differences of squares of reciprocal whole numbers: a critical clue, historically leading to the Bohr atom and beyond ....

The beauty of Fourier's modus operandi is that the fundamental, and all harmonics arise from a single *model* wave (a sine wave). The harmonics are all simply sine waves as well, but modified appropriately in their *frequency, phase,* and *amplitude.* To synthesize any sound wave, by the way, you need only know this numerical data (*frequency* and *amplitude*) of the fundamental, and of each of the higher harmonics. The fact that the human ear doesn't seem to distinguish modifications of phase suggests that Fourier's analysis is particularly germane to an understanding of that perceptive apparatus.

17 A dream of a "Fourier-type" analysis of $P(N)$?

The staircase of primes, $P(N)$, is not periodic in any sense, but still we may dream of roughly following the format set up by Fourier, as described in the previous section, in our search for curves

that closely fit $P(N)$. We might look for a smooth curve that is an initial, relatively close, fit to $P(N)$; let's refer to such an appropriate curve as $F(X)$. The letter "F" could stand for *fit*, or *Fourier*, or *fundamental*, ...) That is, we want our $F(X)$ to play a role analogous to that of Fourier's *fundamental*. We then will be wanting to improve the fit between $F(X)$ and $P(N)$ and correcting $F(X)$ by successively adding more and more *corrective terms* that will play the analogous role of Fourier's *harmonics*. Recall (from our discussion in the previous section) that the harmonics comprising the square wave are simple sine waves calibrated by having a certain *frequency*, *phase*, and *amplitude*. We would hope for something similar here: we would want each of these "harmonics" to be fashioned, somehow, from the "fundamental," that is, from $F(X)$, by recalibrating this initial curve $F(X)$ appropriately. Finally, we might dream of getting a perfect fit for the staircase of primes, if we correct our fundamental $F(X)$ by adding *all* the harmonics to it.

Riemann, in his wonderful 1859 article, did far more than suggest a close-fitting approximation, $R(X)$, to the staircase of primes. He turned the full dream we have just described into a reality. The easiest way of thinking of what extra thing Riemann did is to make use of Fourier's vocabulary. Think of Riemann's smooth curve $R(X)$ as the "$F(X)$" wished for in the previous paragraph; think of it as **Riemann's fundamental** approximation to $P(N)$. Going further, Riemann described an infinite sequence of correction terms to this fundamental approximation; we can refer to them as **Riemann's harmonics**. The harmonics have all of the qualities in our dream above, and if we call $R_k(X)$ Riemann's curve corrected by adding the first k of Riemann's harmonics, then as k tends to infinity, the more and more closely fitting curves $R_k(X)$ converge to provide an *exact fit* for $P(N)$ [see endnote 3].

At this point to get a sense of how elegantly, and rapidly, the successive approximations of Riemann fit the staircase of primes, look at slide 33 which gives us a picture of Riemann's fundamental (labelled $R_0(X)$ there) which is already a pretty good approximation to $P(N)$ even if we are at the high magnification of the miniscule range of the first 100 numbers. Slide 34 shows the comparison between $P(N)$ and $R_5(X)$, Riemann's fundamental corrected by the first five of his harmonics. Slide 35 shows Riemann's fundamental corrected by the first ten of his harmonics. Now watch slides 36-76, which passes from Riemann's fundamental alone, through its corrections via each of the first forty harmonics in turn, to see it rapidly and mysteriously hugging those erratic steps and rises of the staircase of primes. If we went *all the way* we would be getting a perfect fit, but this, already after a mere forty corrections, is pretty good!

18 Fourier Analysis *versus* Fourier Synthesis

If this were the end of the story, a perfectly natural reaction to what we have discussed in the previous section might be: "it is impressive, as an exercise in curve-fitting, but so what?" For it is hardly surprising that an artful analysis of the staircase $P(N)$, or–for that matter–of any graph, might lead to closer and closer approximations to the graph by smooth curves[5]. Let us call that kind of curve-fitting exercise *analyzing the graph*.

Riemann did something much more profound. He put his finger on an elegant mathematical formula– a Rosetta stone– that holds the key to predicting precisely, the entire sequence of harmonics to use in the successive corrections of Riemann's fundamental, in order to perfectly *synthesize the graph* $P(N)$. This Rosetta stone is called the *Riemann zeta function* and it occurs on the very first page of his 1859 manuscript. (There is no need, at this point, to read it exactly, but glance at the displayed equation in slide 77, which is slightly enlarged in slide 78.) We will return to it.

19 Riemann's harmonics

I mentioned that each of Riemann's harmonics is, in some sense, a *recalibration* of Riemann's fundamental, just as–in the case of the square wave–each higher harmonic in Fourier's analysis is a sine wave and hence a recalibration, in some sense, of the fundamental; recalibrated, that is, by having a specific *frequency*, *phase*, and *amplitude*. To specify one of Fourier's harmonics you need only give its frequency, phase, and amplitude.

Similarly, each of Riemann's harmonics will be completely specified if you give two numbers. I'll call them the **essential oscillating frequency** and the **essential error magnitude** of the harmonic.

For what will occur later, the exact interpretation of the essential oscillating frequency is of less interest to us than is the interpretation of the essential error magnitude of a harmonic. Most important, though, is that

If the essential error magnitude of a harmonic is $1/2$ *then the harmonic is square root small; if it is greater than* $1/2$ *then the harmonic is not square root small.*

This statement is what is going to make the connection between our "first view" of the Riemann Hypothesis (section 14) and our "second view" (section 21).

20 Riemann's recipe

The Riemann zeta function, denoted $\zeta(s)$, might in its early stages of development have been better named *The Euler zeta function* after Leonhard Euler who understood some of its importance for the study of numbers. However you name it, the value $\zeta(s)$ of the zeta function for s any real number bigger than 1 is simply the sum of the reciprocals of the s-th powers of all natural numbers, that is:

$$\zeta(s) = 1 + 1/2^s + 1/3^s + 1/4^s + 1/5^s + \dots$$

Euler knew that this was intimately related to the properties of prime numbers because –thanks to the fact that any number is uniquely expressible as a product of prime numbers– Euler expressed the infinite sum above as an infinite product of factors of the form

$$\frac{1}{1 - 1/p^s}$$

where p's run through all prime numbers.

Thinking of $\zeta(s)$ as the infinite sum though, it is clear that $\zeta(s)$ can *never* tote up to the value 0 for any real number s bigger than 1, because $\zeta(s)$ is an infinite sum of positive numbers. But here is where Riemann's attitude towards $\zeta(s)$ changes the game entirely. Riemann saw a way of naturally extending the definition of the function $\zeta(s)$ so that it makes sense not just for real numbers s that are bigger than 1, but for *all* complex numbers s. "Naturally" is the key word in the previous sentence, of course. In this larger domain of all complex numbers, $\zeta(s)$ may take the value 0 for certain complex numbers s. If $s = a + ib$ is such a complex number–that is, if $\zeta(a + ib) = 0$, say that "$a + ib$ is a zero of the zeta-function." Riemann knew, as well, that among the complex numbers $s = x + iy$ where the real part, x, is between 0 and 1 (including, possibly 0 and 1) and y is positive, there were infinitely many zeroes of his zeta-function. Let us call the region of the complex plane consisting of complex numbers with real parts x between 0 and 1 (including, possibly 0 and 1) and imaginary parts y positive, the **positive critical strip**.

Here, then, is Riemann's extraordinary recipe for determining what we called (in the previous section) Riemann's harmonics, and more specifically here is the recipe for determining the *essential*

oscillating frequency and *essential error magnitude* of the harmonics that Riemann obtains to get his *perfect fit* to $P(N)$. Every zero $s = a + ib$ of the zeta-function $\zeta(s)$ in the positive critical strip corresponds to one of Riemann's harmonics, and conversely. The essential error magnitude of the harmonic corresponding to the zero $a + ib$ of $\zeta(s)$ is equal to the real part of the complex number $a + ib$ (i.e., it is equal to a) while the essential oscillating frequency of the harmonic is equal to the imaginary part of $a + ib$ (i.e., it is equal to b). From this data, then, Riemann could *synthesize* the staircase of primes; that is, if he managed to locate all the zeroes of his zeta function that fall in the positive critical strip.

21 The Riemann Hypothesis: second view

At the time of publication of his 1859 article Riemann did manage to compute the first dozen or so zeroes of his zeta function (lying in the positive critical strip). He discovered that they all lined up on the vertical line of complex numbers with real part precisely 1/2. This is astoundingly important for error estimates, because it meant that the harmonics corresponding to those zeroes are all *square root small*, and in particular, the corresponding curves $R_k(X)$ are all square root close to the fundamental $R(X)$. Our second view of the Riemann Hypothesis–perfectly equivalent to the first view–is:

The Hypothesis of Riemann (second view): *All the zeroes of the zeta function in the positive critical strip lie on the vertical line consisting of those complex numbers with real part equal to* 1/2.

See slide 79 for a picture of (a finite piece of) the positive critical strip of complex numbers. The absolute value of the Riemann zeta function is indicated in the strip by color, the darker colors correspond to lower absolute value, the brighter colors to higher, with the exception that the actual zeroes of the zeta function are indicate by white dots, so that they stand out. All the forty odd zeroes depicted on that slide lie on the line predicted by RH.

We now know that the first ten trillion zeroes of the zeta-function lie on the middle vertical line! But does this pattern persist? We will only know that it does for sure once the Riemann Hypothesis is proved. If you want to see tables of the first $100,000$ of these zeroes lovingly calculated to an accuracy within $3 \cdot 10^{-9}$, consult Andrew Odlyzko's tables:
http://www.dtc.umn.edu/~odlyzko/zeta_tables/.

22 Rogue zeroes

What would happen if RH were false, i.e., if a rogue zero of the zeta-function were to be discovered off the middle vertical line in the positive critical strip, as is fantasaically depicted in slide 82? First, the Riemann zeta function enjoys a certain symmetry property that would then guarantee the existence of another rogue zero symmetrically placed in the positive critical strip, as shown in slide 82. But more importantly, Riemann's harmonic that corresponds to the rogue zero to the right of the middle vertical line would not be "square root small." The correction to Riemann's fundamental, then, that takes this (rogue) harmonic into account *would not be square-root close to the initial fundamental,* and (worse!) neither Gauss's curve $G(X)$ nor Riemann's would then be approximations to $P(N)$ with only square-root error.

If RH were false, how bad would the error be between $G(X)$, or between $R(X)$ and $P(N)$? The further to the right a rogue zero lies, the worse the error. For example, if we only knew that the real part of all rogue zeroes are less than some number less than 3/4 we could say that the difference between $R(N)$ and $P(N)$ is at worst $N^{0.75}$ for large N, but we could say nothing sharper.

23 Extreme rogue zeroes: The Prime Number Theorem

What if the rogue zero were at the very rim of the critical strip? Say with real part equal to 1? If there were such a zero, this would be disaster for both Gausss curve and Riemanns curve because it would imply that their accuracy– as a curve that closely fits $P(N)$– could be guaranteed to be no better than their size. But if your error term is as big as what you are trying to measure, you may as well go home!

Happily, there is no such zero. That such extreme rogue zeroes do not exist is the essential content of *The Prime Number Theorem,* the result that we have encountered earlier that provided us with a guarantee that the error term between $G(N)$ and $P(N)$ (and also between $R(N)$ and $P(N)$) is sufficiently small so that the ratios

$$G(N)/P(N)$$

tend to 1 as N gets larger and larger. That this was true for Gauss's curve $G(N)$ was surmised by Gauss himself in the 1790s, but was proved only a century later (independently by Hadamard and De la Valée Poussin).

24 Key and Signature

Even once we prove the Riemann Hypothesis we will be barely at the beginning of this story. What ARE the zeroes of the zeta-function? Are they the key as well as the signature of the staircase of primes? Are they signals to us of a subatomic structure of prime numbers?

That a simple geometric property of these zeroes (lying on a line!) is directly equivalent to such profound (and more difficult to express) regularities among prime numbers suggests that these zeroes and the parade of Riemann's corrections governed by them–when we truly comprehend their message–may have lots more to teach us, may eventually allow us a more powerful understanding of arithmetic. This infinite collection of complex numbers, i.e., the nontrivial zeroes of the Riemann zeta function, plays a role with respect to $P(N)$ rather like the role the *spectrum* of the Hydrogen atom, plays in Fourier's theory. Are the primes themselves no more than an epiphenomenon, behind which there lies, still veiled from us–a yet-to-be-discovered, yet-to-be-hypothesized, profound conceptual key to their perplexing orneriness? Are the many innocently posed, yet unanswered, phenomenological questions about numbers–such as in the ones listed earlier– waiting for our discovery of this deeper level of arithmetic? Or for layers deeper still? These are not completely idle

thoughts, for a tantalizing analogy relates the number theory we have been discussing to an already established branch of mathematics–due, largely, to the work of Alexander Grothendieck, and Pierre Deligne–where the corresponding analogue of Riemann's hypothesis has indeed been proved...

Some technical end-notes.

(1) Gauss's guess that I denote $G(X)$ in the text above, and is usually denoted $\mathrm{Li}(x)$, is

$$\mathrm{Li}(x) = \int_2^\infty dx/\log(x),$$

while Riemann's guess is

$$R(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} \mathrm{Li}(x^{\frac{1}{n}}),$$

where $\mu(n)$ is the Moebius function.

(2) If $f(x)$ and $g(x)$ are real-valued functions of a real variable x such that for any $\epsilon > 0$ both of them take their values between $x^{1-\epsilon}$ and $x^{1+\epsilon}$ for x sufficiently large, then say that $f(x)$ and $g(x)$ are **approximations of one another with square-root accuracy** if, for any positive ϵ the absolute values of their difference is less than $x^{\frac{1}{2}+\epsilon}$ for x sufficiently large. The functions $\mathrm{Li}(x)$ and $R(x)$ of end-note **(1)** are approximations of one another with square-root accuracy.

(3) For people who know the technical theory behind all this, the text has not just been sketching a *mere analogy* between Riemann's ideas and Fourier's: the relation between the two viewpoints is very close! Riemann worked with a function easily related to $P(N)$, that he called $\psi(X)$, and, in effect provided a (strictly) Fourier analysis of $\psi(X)$, viewed as a function on the multiplicative group of positive real numbers. The minor miracle is that the Fourier transform of $\psi(X)$ is a discretely supported distribution, its support being given–essentially– by the multiplicative characters

$$X \mapsto X^{\rho}$$

where ρ runs through the zeroes of the zeta-function on the full critical strip. Riemann then *retro-fitted* this information as a way of obtaining an explicit formula for what I denote in the text as $P(N)$. Riemann's harmonic–as I refer to it in the text–corresponding to a zero ρ of the zeta-function in the positive critical strip is given by the equation

$$-R(X^{\rho}) - R(X^{\bar{\rho}}),$$

so the successive approximations of Riemann denoted $R_k(X)$ are given by the equations

$$R_k(X) := R(X) - \sum_{j=1}^{k} (R(X^{\rho_j}) + R(X^{\bar{\rho}_j})).$$

where the ρ_j run through the first k zeroes of the Riemann zeta-function in the positive critical strip. If you happen to wonder how I propose to order these zeroes if RH is false, the following prescription will do: order them in terms of their imaginary part, and in the unlikely situation that there is more than one zero with the same imaginary part, order zeroes of the same imaginary part by their real parts, going from right to left.

References

typo: page 18, messed up close parenthesis in second equation.