Primes: An Introduction, From Euclid's Proof to the Forest Floor

Ian Donaldson
Independent Researcher
(Dated: August 5, 2026)

This is an introduction to prime numbers, written for anyone curious about them regardless of any other interest. It begins with the first people on record who treated primality as a structurally special property worth proving things about, rather than a label applied to individual numbers one at a time. It looks at a real, still-debated case where prime numbers appear to matter to survival in the natural world, the periodical cicada. It works through the classical connection between primes and which regular polygons can be built with a compass and straightedge, and the visual pattern in the Ulam spiral. It closes with two genuine pieces of mathematical machinery, the Prime Number Theorem and Stirling's approximation, developed carefully enough to actually use rather than only quoted.

I. INTRODUCTION

Prime numbers, integers greater than one divisible only by one and themselves, were not first noticed as a curiosity. They were first noticed as a structural fact worth proving something about. Euclid, working in Alexandria around 300 BCE, gave a proof that there are infinitely many of them, in Book IX of the *Elements* [1]. The proof is short enough to give here in full, because it is worth seeing exactly as it was first constructed, not paraphrased.

Suppose there are only finitely many primes, $p_1, p_2, \ldots, p_n$. Form the number $N = p_1 p_2 \cdots p_n + 1$. Every one of p_1 through p_n leaves a remainder of exactly 1 when divided into N, so none of them divides N. But every integer greater than 1 has at least one prime factor, so N has a prime factor, and that factor is not on the original list. The list was assumed to contain every prime, and it does not. No finite list can contain every prime, so there are infinitely many.

A near contemporary, Eratosthenes of Cyrene, gave the first systematic method for finding every prime up to a given bound: write down every integer from 2 to that bound, then repeatedly cross out the multiples of the smallest number not yet crossed out, starting from 2 [2]. What survives is exactly the primes. This method, the sieve of Eratosthenes, is still the fastest simple way to generate a real list of primes today, and every sieve computation in this paper uses it directly.

II. A PATTERN FOUND IN NATURE

Most of what follows in this paper is about pure structure. One case is worth opening with instead, because it suggests prime numbers might matter outside mathematics entirely, in a place nobody was looking for them: the life cycle of an insect.

The genus *Magicicada*, found in the eastern United States, contains species that spend either exactly 13 or exactly 17 years underground as juveniles before emerging together, all at once, to breed [3]. Both 13 and 17 are prime. This has a real, actively studied explanation,

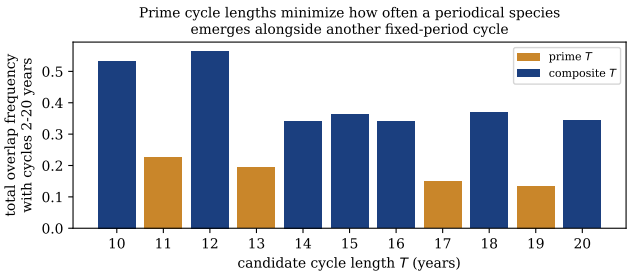

FIG. 1. Total overlap frequency, $\sum_k 1/\mathrm{lcm}(T,k)$ for $k = 2, \ldots, 20$, $k \neq T$, against candidate cycle length T. Prime values of T (amber) synchronize with other fixed cycle lengths less often than any composite value tested.

though it remains a hypothesis rather than settled fact: a population that emerges on a prime-numbered cycle synchronizes with a competing or predatory cycle of any other fixed length far less often than a population on a composite-numbered cycle would [4, 5].

The reason is a fact about least common multiples. Two cycles of length a and b coincide every $\mathrm{lcm}(a,b)$ years. When a is prime, $\mathrm{lcm}(a,b) = ab$ for any b that is not itself a multiple of a, the largest value it can possibly take, so a prime-cycle population meets any other fixed-period population as rarely as arithmetically possible. Figure 1 makes this concrete: summing $1/\mathrm{lcm}(T,k)$ over candidate competing cycle lengths k from 2 to 20, as a direct measure of how often a cycle of length T would fall into step with something else, prime values of T score lowest, and by a wide margin.

This does not prove why real cicadas settled on 13 and 17 specifically rather than, say, 19; real evolutionary history involves predator populations, climate history, and genetic drift that this simple count does not capture, and more than one competing hypothesis exists in the literature [4, 5]. What the figure does establish honestly is the underlying arithmetic fact the biological hypothesis rests on: prime cycle lengths are a real, provable local minimum of synchronization frequency, not a loose analogy.

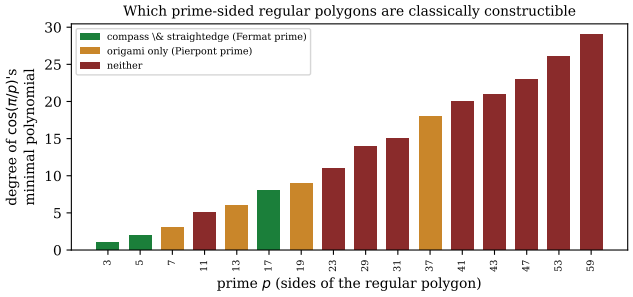

FIG. 2. Every prime $p < 60$, classified by whether the regular p-gon is constructible with compass and straightedge (green), constructible by the more powerful method of paper folding, or origami, which can additionally solve cubic equations (amber), or neither (red). Bar height is the degree of the minimal polynomial of $\cos(\pi/p)$, which is $(p-1)/2$ for every prime p.

III. PRIMES AND GEOMETRY: WHICH POLYGONS CAN BE BUILT

A second place primes turn up, purely inside mathematics this time, is in deciding which regular polygons can be constructed with an unmarked straightedge and a compass, the classical Greek tools. Gauss proved in 1796, at nineteen years old, that the regular 17-gon is constructible, the first genuine advance on this question since antiquity [13]. The complete answer, due to Gauss and Wantzel together, is that a regular n-gon is constructible exactly when $n = 2^k p_1 p_2 \cdots p_m$, where the p_i are distinct *Fermat primes*, primes of the form $2^{2^j} + 1$ [14]. Only five are known to exist at all: 3, 5, 17, 257, and 65,537.

The connection to primality runs through the vertices of these polygons themselves. The n-th roots of unity, the n complex numbers solving $z^n = 1$, sit at n evenly spaced points on the unit circle, and the polynomial whose roots are exactly the roots of unity that are not shared with any smaller divisor of n is called the n-th cyclotomic polynomial. For a prime index p, this polynomial has degree $p - 1$, and the real coordinate of a single vertex, $\cos(2\pi/p)$, satisfies a polynomial of exactly half that degree, $(p-1)/2$. Whether that degree happens to be a power of two decides constructibility outright, which is exactly why only five primes, out of infinitely many, ever qualify. Figure 2 shows this directly for every prime under 60.

IV. THE ULAM SPIRAL

A different kind of prime pattern is visual rather than arithmetic. Write the positive integers along a square spiral, starting at the center, and mark only the primes. Stanislaw Ulam noticed, doodling during a conference talk in 1963, that the marked primes are not spread evenly across the resulting grid [15]. They cluster preferentially along certain diagonal lines, corresponding to quadratic polynomials in the spiral's own winding coordinate, such as $n^2 + n + 41$, a polynomial already known in the eighteenth century to produce an unusually long run of primes for consecutive small n. Figure 3 reproduces the pattern directly, from the actual spiral and the actual primes up to 160,802, not an illustrative sketch. The clustering is real and reproducible, and it is explained, though not fully proved in general, by the Hardy–Littlewood conjectures on how densely a given quadratic polynomial's values are populated by primes [16].

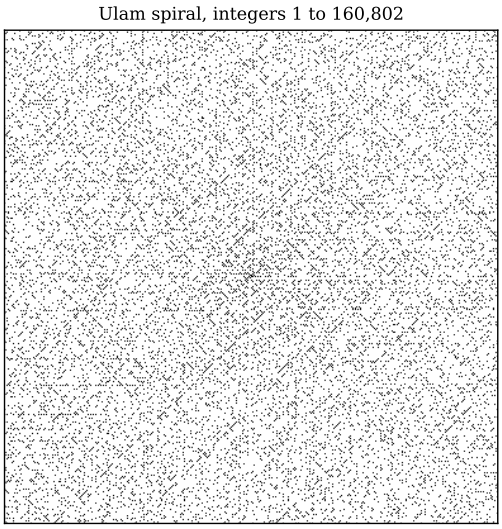

FIG. 3. The Ulam spiral, generated directly: integers 1 through 160,802 on a square spiral, primes marked black. The diagonal clustering reflects real, unusually prime-rich quadratic polynomials running along those diagonals.

V. THE GAPS BETWEEN PRIMES

The Prime Number Theorem describes primes on average. It says nothing about how irregularly they actually sit next to each other, and that irregularity is itself real and provable. Bertrand's postulate, conjectured in 1845 and proved by Chebyshev in 1852, guarantees at least one prime strictly between n and $2n$ for every $n > 1$ [20], a strong global regularity statement. The local picture is the opposite. For any integer n, the n consecutive integers $n! + 2, n! + 3, \ldots, n! + n + 1$ are all composite, since k divides $n! + k$ for every k from 2 to $n + 1$. Arbitrarily long gaps between consecutive primes certainly exist, and they can be forced into existence at a location fixed in advance.

At the opposite extreme sits the twin prime conjecture: that infinitely many pairs $(p, p+2)$ exist, the smallest possible gap repeating forever. It remains unproved, despite overwhelming computational evidence and twin

pairs found into the extremely large range. Real progress exists short of a full proof. Yitang Zhang proved in 2013 that some gap size below 70,000,000 recurs infinitely often, the first finite bound of any kind on this question in history [6]. The Polymath8 collaboration, and independently James Maynard, then reduced that bound to 246 using different sieve methods [7, 8]. The gap between 246 and the conjectured value of 2 remains open.

VI. A SECOND ANCIENT PATTERN: PERFECT NUMBERS

Euclid's own book contains a second prime-related result, less famous than the infinitude proof but proved in the same work: a recipe for building what the Greeks called a perfect number, an integer equal to the sum of its own proper divisors. Book IX, Proposition 36, shows that whenever $2^p - 1$ is prime, the number

$$N = 2^{p-1}\,(2^p - 1)$$

is perfect [1]. Take $p = 2$: $2^2 - 1 = 3$ is prime, and $N = 2(3) = 6$; its proper divisors are 1, 2, and 3, which sum to 6 exactly. Take $p = 3$: $2^3 - 1 = 7$ is prime, and $N = 4(7) = 28$; its proper divisors, $1, 2, 4, 7, 14$, sum to 28.

A prime of the specific form $2^p - 1$ is now called a Mersenne prime, after Marin Mersenne, who studied them in 1644, nineteen centuries after Euclid's original construction [9]. The exponent p itself must be prime, though this is far from sufficient: $2^{11}-1 = 2047 = 23\times89$ shows a prime exponent can still fail. Mersenne primes are the direct target of the largest ongoing distributed-computing prime search running today, because an efficient primality test specific to this exact form, the Lucas–Lehmer test, exists [10], and finding one immediately hands over a new, explicitly nameable perfect number through Euclid's two-thousand-year-old formula. Euler proved, in the eighteenth century, that Euclid's construction accounts for every even perfect number without exception, closing the other half of the question Euclid had opened [11]. Whether even a single odd perfect number exists is, as of this writing, still completely open, checked computationally into the extremely large range and proved in no case at all.

VII. HOW MANY PRIMES ARE THERE, REALLY

Euclid's proof establishes that the primes never run out. It says nothing about how common they are at a given size. Let $\pi(x)$ denote the count of primes up to x. The Prime Number Theorem, proved independently by Hadamard and de la Vallée Poussin in 1896, states

$$\pi(x) \sim \frac{x}{\ln x}.$$

This is a genuinely logarithmic law: primes thin out as x grows, at a rate set by the natural logarithm of x, not by any power of x. Verified here directly by exhaustive sieve up to $x = 200{,}000$, the true count is $\pi(200{,}000) = 17{,}984$, against the simple approximation $x/\ln x \approx 16{,}385$; the tighter logarithmic-integral estimate $\mathrm{Li}(x) = \int_2^x dt/\ln t$ gives 18,036, visibly closer on both sides.

VIII. A DEEPER METHOD: STIRLING AND CHEBYSHEV

The last piece of this introduction is a genuine tool, not just a fact: a way to actually estimate how many primes must exist in a given range, developed from first principles rather than quoted.

A. Stirling's Approximation

James Stirling published, in 1730, an asymptotic formula for the factorial [17]:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

This is not a convergent series and does not become exact as $n \to \infty$ in the ordinary sense; it is an asymptotic expansion, useful when truncated, and the way to actually derive it connects directly back to Taylor series. The Euler–Maclaurin formula, developed independently by Euler in 1735 and Maclaurin in 1742, converts a discrete sum such as $\ln 1+\ln 2+\cdots+\ln n$ into an integral plus a series of correction terms built from the derivatives of the summand at its endpoints [18, 19]. The coefficients in that correction series are the Bernoulli numbers, and the Bernoulli numbers are themselves defined as the Taylor coefficients of the generating function $t/(e^t-1)$. Stirling's formula is what falls out of applying this machinery to $\ln(n!)$.

The approximation converges quickly despite being only asymptotic. Direct computation gives $5! = 120$ against Stirling's estimate of 118.02, an error of about 1.6%; by $n = 20$ the relative error has fallen to about 0.4%, shown in Figure 4.

B. Chebyshev's Bound

Stirling's formula becomes a real tool for counting primes through a clever choice of object to estimate: the central binomial coefficient $\binom{2n}{n} = (2n)!/(n!)^2$. Legendre's formula gives the exact power of a prime p dividing $n!$ as $\sum_{i\geq1}\lfloor n/p^i\rfloor$, and applying it to $\binom{2n}{n}$ shows that every prime strictly between n and $2n$ divides this coefficient to exactly the first power, no more and no less. This is not an approximation. Direct computation confirms it exactly: $\binom{20}{10} = 184{,}756 = 2^2 \times 11 \times 13 \times 17 \times 19$, and

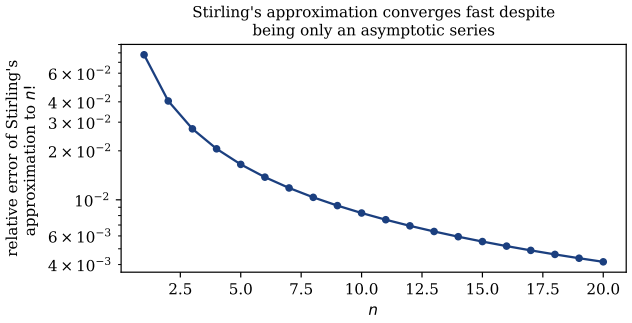

FIG. 4. Relative error of Stirling's approximation to $n!$, computed directly against the exact factorial, for $n = 1$ through 20.

each of the primes between 10 and 20, namely 11, 13, 17, and 19, appears with exponent exactly 1.

Pairing this exact arithmetic fact with Stirling's estimate of $\binom{2n}{n}$'s overall size gives a real, two-sided bound on how many primes must lie between n and $2n$, the method Chebyshev used in 1852 to prove the first rigorous density bounds on the primes, decades before the full Prime Number Theorem was proved [20]. The same argument, cleaned up, is the heart of Erdős's celebrated one-page 1932 proof that a prime always exists between n and $2n$ for every $n > 1$, a statement first conjectured by Bertrand in 1845 [21].

IX. WHAT IS STILL NOT KNOWN

The Prime Number Theorem describes the average behavior of $\pi(x)$ well, but says nothing about exactly how far a real count can wander from that average at any particular x. The sharpest possible version of that statement is the Riemann Hypothesis, proposed by Bernhard Riemann in 1859 [12]: every non-trivial zero of the Riemann zeta function, $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$, extended to the complex plane, has real part exactly $\frac{1}{2}$. This is equivalent to the tightest error bound the Prime Number Theorem could possibly have. It has been checked numerically for the first several trillion zeros without a single exception, and it remains, over one hundred and sixty years after Riemann wrote it down, unproved.

X. CONCLUSION

Prime numbers were first treated as a structural fact, not a curiosity, by Euclid, and the two-thousand-year-old proof that they never run out still reads exactly as it was first given. They appear to matter in the real, evolved life cycle of an insect, through a provable fact about least common multiples rather than a loose numerical coincidence. They govern, completely and exactly, which regular polygons can be built by hand, through their connection to the roots of unity. They cluster visibly along diagonals in a spiral doodled during a conference talk. And a genuine piece of eighteenth-century calculus, Stirling's approximation, built from the same Taylor-series machinery that underlies the Euler–Maclaurin formula, turns into a real tool for proving how many of them there must be. None of this required assuming anything beyond what could be checked directly.

[1] Euclid, *Elements*, Book IX, Proposition 20 (c. 300 BCE).

[2] Nicomachus of Gerasa, *Introduction to Arithmetic* (c. 100 CE), the earliest surviving description of the sieve of Eratosthenes.

[3] K. S. Williams and C. Simon, "The Ecology, Behavior, and Evolution of Periodical Cicadas," Annual Review of Entomology **40**, 269 (1995).

[4] J. Yoshimura, "The evolutionary origins of periodical cicadas during ice ages," The American Naturalist **149**, 112 (1997).

[5] E. Goles, O. Schulz, and M. Markus, "Prime number selection of cycles in a predator-prey model," Complexity **6**, 33 (2001).

[6] Y. Zhang, "Bounded gaps between primes," Annals of Mathematics **179**, 1121 (2014).

[7] D. H. J. Polymath, "Variants of the Selberg sieve, and bounded intervals containing many primes," Research in the Mathematical Sciences **1**, Art. 12 (2014).

[8] J. Maynard, "Small gaps between primes," Annals of Mathematics **181**, 383 (2015).

[9] M. Mersenne, *Cogitata Physico-Mathematica* (Paris, 1644).

[10] D. H. Lehmer, "On Lucas's Test for the Primality of Mersenne's Numbers," Journal of the London Mathematical Society **10**, 162 (1935).

[11] L. Euler, *Tractatus de numerorum doctrina*, posthumously published, Commentationes Arithmeticae **2** (1849).

[12] B. Riemann, "Über die Anzahl der Primzahlen unter einer gegebenen Grösse," Monatsberichte der Berliner Akademie (1859).

[13] C. F. Gauss, *Disquisitiones Arithmeticae* (Gerhard Fleischer, Leipzig, 1801).

[14] P. L. Wantzel, "Recherches sur les moyens de reconnaître si un problème de géométrie peut se résoudre avec la règle et le compas," Journal de Mathématiques Pures et Appliquées **2**, 366 (1837).

[15] M. L. Stein, S. M. Ulam, and M. B. Wells, "A Visual Display of Some Properties of the Distribution of Primes," The American Mathematical Monthly **71**, 516 (1964).

[16] G. H. Hardy and J. E. Littlewood, "Some problems of 'Partitio numerorum' III: On the expression of a number as a sum of primes," Acta Mathematica **44**, 1 (1923).
[17] J. Stirling, *Methodus Differentialis* (London, 1730).
[18] L. Euler, "Methodus generalis summandi progressiones," Commentarii Academiae Scientiarum Petropolitanae **6**, 68 (1735).
[19] C. Maclaurin, *A Treatise of Fluxions* (Edinburgh, 1742).
[20] P. L. Chebyshev, "Mémoire sur les nombres premiers," Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg **7**, 15 (1852).
[21] P. Erdős, "Beweis eines Satzes von Tschebyschef," Acta Scientiarum Mathematicarum **5**, 194 (1932).