

# The work of James Maynard

Kannan Soundararajan

## Abstract

We give a brief account of some of the most spectacular results established by James Maynard, for which he has been awarded the Fields Medal.

## Mathematics Subject Classification 2020

Primary 11N05; Secondary 11N32, 11N35, 11J83

## Keywords

distribution of primes, sieve methods, metric Diophantine approximation

James Maynard has established several spectacular results in analytic number theory. While the proofs of these results involve many deep ideas, their statements are remarkable for their simplicity and elegance. To illustrate, we state two such striking results of Maynard concerning prime numbers, before setting them in context.

**Theorem 1.** (Maynard [32]) *For each natural number  $m \geq 2$ , there exists a positive integer  $C(m)$  with the following property: There are infinitely many natural numbers  $n$  such that the interval  $[n, n + C(m)]$  contains at least  $m$  prime numbers.*

**Theorem 2.** (Maynard [36]) *There are infinitely many prime numbers  $p$  whose decimal representation does not contain the digit 7.*

**Background.** To place these results in context, recall that the prime number theorem gives an asymptotic for  $\pi(x)$ , the number of primes below  $x$ ; namely,

$$\pi(x) \sim \text{li}(x) = \int_0^x \frac{dt}{\log t}.$$

We may think of this asymptotic as roughly saying that the “chance” of a number  $n$  being prime is about  $1/\log n$ . One overarching theme in analytic number theory may be formulated as asking in what ways does the sequence of primes resemble, or differ from, a random sequence of integers with each integer  $n \geq 3$  chosen independently to be in the random sequence with probability  $1/\log n$  (this is also known as the Cramér model). One obvious difference is that all primes larger than 2 must be odd, whereas a random sequence would surely contain many even numbers. But if we could account for divisibility by small primes (such as 2 in our example), would a modified random model describe accurately the behavior of prime numbers?

There are many ways in which we could try to make this theme precise. For instance, the Riemann hypothesis predicts that  $|\pi(x) - \text{li}(x)|$  is bounded by  $C(\epsilon)x^{\frac{1}{2}+\epsilon}$  for any  $\epsilon > 0$  and some constant  $C(\epsilon)$ . Fluctuations of size about  $\sqrt{x}$  are indeed what one would expect if we select random sets of integers with  $n \geq 3$  included in the set with probability  $1/\log n$ . Thus the Riemann hypothesis is, at a crude level, consistent with a random model of primes, although if we inspect the error term  $\pi(x) - \text{li}(x)$  in finer detail then the influence of zeros of  $\zeta(s)$  would be visible, and such features would deviate (in small but significant ways) from the random model.

At the 1912 ICM, Landau posed four “unattackable” problems on primes: (i) the Goldbach problem that every even integer larger than 2 is the sum of two primes, (ii) the twin prime problem that there are infinitely many prime pairs  $n$  and  $n + 2$ , (iii) there is always a prime between two consecutive squares, and (iv) there are infinitely many primes of the form  $n^2 + 1$ . All four problems remain open today, and all statements are exactly what one would expect for random sequences. For example, the Cramér model would suggest that the chance that  $n$  and  $n + 2$  are both “prime” is about  $1/\log n \times 1/\log(n + 2)$ , which would predict about  $x/(\log x)^2$  twin primes up to  $x$ . Of course some care is needed, since the same prediction could be made for  $n$  and  $n + 1$  being prime, and we will address this soon. Similarly, we may expect that an even number  $N$  may have about  $N/(\log N)^2$  representations as a sum of two

primes, making the Goldbach conjecture very plausible, and related arguments suggest the last two Landau problems as well.

For the third Landau problem on the number of primes between  $n^2$  and  $(n+1)^2$ , the random model already predicts what we believe to be the right answer — namely, there should be about  $(2n+1)/\log(n^2) \approx n/\log n$  primes in this interval. For the other three problems, some modification must be made to the Cramér model, to take into account the deterministic features of these problems with respect to divisibility by small primes. Precise conjectures for these problems were first made by Hardy and Littlewood motivated by their work on the circle method. These conjectures are widely believed to be true, and are supported by extensive heuristic and numerical evidence. For instance, Hardy and Littlewood formulated the following conjecture for the number of twin primes below  $x$ :

$$\#\{n \leq x : n, n+2 \text{ both prime}\} \sim \mathfrak{S}(\{0, 2\}) \int_2^x \frac{dt}{(\log t)^2}.$$

Here  $\int_2^x dt/(\log t)^2$  is asymptotically  $x/(\log x)^2$ , and corresponds to the prediction of the Cramér model, while  $\mathfrak{S}(\{0, 2\})$ , known as the *singular series*, is a correction factor

$$\mathfrak{S}(\{0, 2\}) = 2 \prod_{p \geq 3} \left(1 - \frac{2}{p}\right) \left(1 - \frac{1}{p}\right)^{-2} = 1.32 \dots$$

The constant  $\mathfrak{S}(\{0, 2\})$  has a compelling probabilistic interpretation: it is a product over all primes  $p$  (the first factor 2 corresponds to the prime  $p = 2$ ), with the factor at  $p$  keeping track of the ratio between the chance that  $n$  and  $n+2$  are not divisible by  $p$ , and the chance that two random numbers are not divisible by  $p$ . Thus, for  $p = 2$ , the chance that  $n$  and  $n+2$  are both not divisible by 2 is  $(1 - 1/2)$  ( $n$  must be odd), while the chance that two random numbers are both not divisible by 2 is  $(1 - 1/2)^2 = 1/4$ ; the ratio of these chances gives the correction factor 2. For primes  $p \geq 3$ , the chance that  $n$  and  $n+2$  are both not divisible by  $p$  is  $(1 - 2/p)$  whereas the chance that two random numbers are both not divisible by  $p$  is  $(1 - 1/p)^2$ , and we see the corresponding correction factor in the definition of  $\mathfrak{S}(\{0, 2\})$ .

Similar conjectures can be made for the binary Goldbach problem, or for the number of primes of the form  $n^2 + 1$ , modifying and correcting the naive predictions of the Cramér model. To illustrate, we give a generalization of the conjecture for twin primes for counting prime  $k$ -tuples: given distinct integers  $h_1, h_2, \dots, h_k$ , for large  $x$  how many integers  $n \leq x$  are there with  $n + h_1, \dots, n + h_k$  all being prime. Here the Hardy–Littlewood conjecture predicts that

$$\#\{n \leq x : n + h_1, \dots, n + h_k \text{ all prime}\} \sim \mathfrak{S}(\{h_1, \dots, h_k\}) \int_2^x \frac{dt}{(\log t)^k} \quad (1)$$

where, with  $\mathcal{H} = \{h_1, \dots, h_k\}$ ,

$$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{\nu(\mathcal{H}, p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-k}, \quad (2)$$

and  $\nu(\mathcal{H}, p)$  denotes the number of distinct residue classes occupied by the set  $\mathcal{H}$  viewed mod  $p$ . Since  $\nu(\mathcal{H}, p) = k$  if  $p$  is larger than  $\max |h_i - h_j|$ , the product defining  $\mathfrak{S}(\mathcal{H})$  converges absolutely to a non-negative real number, and it equals zero only if  $\nu(\mathcal{H}, p) =$

$p$  for some prime  $p$ . If  $v(\mathcal{H}, p) = p$ , then for any integer  $n$  at least one of the numbers  $n + h_1, \dots, n + h_k$  would be a multiple of  $p$ , and therefore there can be only finitely many integers  $n$  with  $n + h_1, \dots, n + h_k$  all being prime; for example this is what happens if we ask for  $n$  and  $n + 1$  to be prime, or  $n, n + 2, n + 4$  all to be prime. When there is no such divisibility obstruction to  $n + h_1, \dots, n + h_k$  all being prime, the Hardy–Littlewood conjecture predicts a rich supply of such prime  $k$ -tuples. This is perhaps the most central question in prime number theory, and remains open in any situation where  $\mathfrak{S}(\mathcal{H})$  is non-zero.

**Sieve theory.** We have described quickly some of the main motivating questions in the theory of primes. One main source of progress towards these questions is *sieve theory*, and a large part of Maynard’s work lies broadly in this area. A typical problem in sieve theory is to bound the size of sets of integers  $\mathcal{A}$  whose elements are constrained to omit  $v(p)$  given residue classes mod  $p$  for primes  $p$ . For instance the twin prime problem is of this form, as we seek to find integers  $n$  that are neither  $0$  nor  $-2 \pmod{p}$  for all primes  $p \leq \sqrt{n+2}$  (so that  $n$  and  $n + 2$  would both be prime). In great generality sieve methods can produce upper bounds of the conjectured order of magnitude; for example, one can show that the number of twin primes up to  $x$  is no more than 4 times the conjectured Hardy–Littlewood asymptotic. Producing corresponding lower bounds has proved to be a much harder problem, but sieve methods have led to striking partial results such as Chen’s theorem that there are many primes  $p$  for which  $p + 2$  has at most two prime factors, or Iwaniec’s theorem that there are many  $n$  for which  $n^2 + 1$  has at most two prime factors. For a comprehensive treatment of the subject, see [14].

Chen’s theorem and Iwaniec’s theorem exhibit a limitation of traditional sieve methods, known as the *parity problem*, which often prevents us from knowing the parity of elements left unsieved, and thus from producing prime numbers. But in some special cases, sieve methods in conjunction with other analytic input have produced prime numbers. For instance, for large  $x$  Baker, Harman, and Pintz [2] showed that the interval  $[x, x + x^\theta]$  contains at least  $cx^\theta/\log x$  primes, where  $c > 0$  is a constant and  $\theta = 0.525$ ; the Landau problem of producing primes between consecutive squares corresponds to intervals with  $\theta = \frac{1}{2}$ . Another spectacular example is due to Friedlander and Iwaniec [13] who established an asymptotic formula for the number of primes up to  $x$  that may be written as  $n^2 + m^4$ , an approximation to the Landau problem of primes of the form  $n^2 + 1$ . A closely related result of Heath-Brown and Li [25] produces an asymptotic formula for primes of the form  $n^2 + p^4$ , where  $p$  is prime. Yet another beautiful result due to Heath-Brown [24] establishes an asymptotic formula for the number of primes below  $x$  of the form  $n^3 + 2m^3$  with  $m, n \in \mathbb{N}$ . Heath-Brown’s result may be viewed as an approximation to the problem of producing primes of the form  $n^3 + 2$ , but before his work it was not even known if there are infinitely many primes that are the sum of three cubes of natural numbers! A crucial feature of these results is that they deal with primes represented by specializations of *norm forms*. The Friedlander–Iwaniec result is concerned with the norm form  $x^2 + y^2 = N(x + iy)$  associated to the field  $\mathbb{Q}(i)$ , and specializing  $y$  to be a square; Heath-Brown’s result is concerned with the norm form  $N(x + y\alpha + z\alpha^2)$  taking the norm over the field  $\mathbb{Q}(\alpha)$  with  $\alpha = 2^{\frac{1}{3}}$ , and specializing  $z$  to be 0. The results of Friedlander–

Iwaniec and Heath-Brown gave the first examples of thin sequences (in the sense that the number of integers below  $X$  in the sequence is  $\leq X^{1-\delta}$  for some  $\delta > 0$ ) of polynomial values in two or more variables that represent infinitely many primes; no example is known of a polynomial in 1 variable of degree more than 1 that represents infinitely many primes.

Maynard's work [40] gives a substantial generalization of Heath-Brown's approach, and produces many further examples of thin sequences of polynomial values in many variables that represent primes. Consider an algebraic number  $\omega \in \mathbb{C}$  of degree  $n$ , and let  $K$  denote the field  $\mathbb{Q}(\omega)$ . We can associate to this the norm form  $N(\sum_{i=1}^n x_i \omega^{i-1})$ , which is a homogeneous polynomial of degree  $n$  in the variables  $x_1, \dots, x_n$ . A thin polynomial in many variables would be obtained by specializing some of the variables in this norm form to be zero; say, we set  $x_{n-k+1}, \dots, x_n = 0$ , and the number integers below  $x$  represented by such an incomplete norm form would be about  $x^{1-k/n}$ . In the range  $n \geq 4k$ , Maynard establishes an asymptotic formula for the number of primes represented by such an incomplete norm form, when the variables  $x_1, \dots, x_{n-k}$  take integer values in the range  $[1, X]$ .

**The circle method.** Apart from sieve theory, another important source of progress towards problems on primes is the *circle method*, which as we already mentioned formed the original motivation for Hardy and Littlewood in formulating their conjectures. To illustrate, consider the Goldbach problem of representing an even integer  $N$  as a sum of two primes. Using Fourier analysis, the number of such representations of  $N$  may be written as

$$r(N) = \int_0^1 S(\alpha)^2 e^{-2\pi i N \alpha} d\alpha, \quad \text{where} \quad S(\alpha) = \sum_{p \leq N} e^{2\pi i p \alpha}. \quad (3)$$

The idea in the circle method is that generating functions such as  $S(\alpha)$  above tend to be large near rational numbers with small denominator (the *major arcs*) and small away from them (the *minor arcs*).

While the circle method has not been able to tackle the binary Goldbach problem or the problem of twin primes, it has been extremely effective in problems where there is a bit more freedom. For instance, the ternary Goldbach problem asks to represent odd numbers as a sum of three primes, and there is one extra variable to play with here. Vinogradov famously used the circle method to show that all large odd numbers are the sum of three primes, and Helfgott [26] has extended this to show that all odd numbers larger than 5 may be so represented. Here we may mention an impressive result of Matomäki, Maynard, and Shao [30] which shows that large odd numbers  $n$  may be expressed as  $p_1 + p_2 + p_3$ , where all three primes  $p_i$  lie in a short interval  $[n/3 - n^\theta, n/3 + n^\theta]$  for any  $\theta > 11/20$ . We mentioned earlier the work of Baker, Harman and Pintz [2] showing the existence of primes in short intervals  $[x, x + x^{0.525}]$ , and the work of [30] is remarkable in solving the ternary Goldbach problem using primes in only slightly longer intervals.

A second example of what it might mean to have an extra degree of freedom is the Green–Tao theorem that the primes contain arbitrarily long arithmetic progressions  $n, n + d, \dots, n + (k - 1)d$ . The Hardy–Littlewood conjecture would predict a stronger “one dimensional” version of such a result with specified choices for the common difference  $d$ ; for instance, there should be infinitely many  $k$ -tuples primes of the form  $n, n + k!, n + 2 \cdot k!$ ,

$\dots, n + (k - 1) \cdot k!$ . The work of Green, Tao, and Ziegler [19–21] may be thought of as a far-reaching generalization of the circle method, obtaining asymptotic formulae for the number of prime solutions to linear systems with at “least two degrees of freedom.”

Maynard’s beautiful result on primes with missing digits (Theorem 2 stated above) is a rare occasion where the circle method can be used to solve a binary problem. Let  $\mathcal{M}$  denote the set of natural numbers with no 7 in their decimal expansion (naturally one could omit any other digit instead of 7). The number of integers in  $\mathcal{M}$  up to  $N$  is about  $N^{\log 9/\log 10} = N^{1-\delta}$  with  $\delta = 0.046\dots$ , so that  $\mathcal{M}$  is a thin set making the problem of finding primes in it a challenge. Before Maynard’s work, Dartyge and Mauduit [7, 8] had used sieve theory to show that  $\mathcal{M}$  contains integers with at most two prime factors. To count the number of primes in  $\mathcal{M}$  up to  $N$ , we use Fourier analysis writing this as

$$\sum_{\substack{p \leq N \\ p \in \mathcal{M}}} 1 = \int_0^1 S(\alpha) M(-\alpha) d\alpha,$$

where  $S(\alpha)$  is the exponential sum over primes defined in (3), and  $M(\alpha) = \sum_{m \leq N, m \in \mathcal{M}} e^{2\pi i \alpha m}$  is the corresponding exponential sum over the set  $\mathcal{M}$ . Usually such a binary problem is hopeless to attack via the circle method — the reason being that even most optimistically we may only expect “square-root cancellation” in the exponential sums  $S(\alpha)$  and  $M(-\alpha)$  for generic  $\alpha$ , and even that would produce an integrand of size  $N^{\frac{1}{2}} \times N^{\frac{1}{2}(1-\delta)}$ , which is bigger than the expected main term of size about  $N^{1-\delta}/\log N$ . A crucial feature in this problem is that the set  $\mathcal{M}$  has a very convenient structure which results in the exponential sum  $M(\alpha)$  often being unusually small. For instance, Maynard shows that its  $L^1$ -norm satisfies

$$\int_0^1 |M(\alpha)| d\alpha \ll N^{0.32},$$

with the key point being that the exponent 0.32 is smaller even than  $(1 - \delta)/2$ , which is the optimistic square-root cancellation that we mentioned. Such estimates raise the hope of being able to attack Theorem 2, and the main idea can be seen transparently in Maynard’s expository article [35], where he proves an easier version of Theorem 2 treating primes missing a digit in base  $b$  with  $b$  sufficiently large. The set of integers up to  $N$  missing a digit in base  $b$  has size about  $N^{\log(b-1)/\log b}$ , and so the problem becomes easier as the base  $b$  gets larger. Getting the base down to 10 turns out to be a fiendishly difficult problem, and is arguably more significant psychologically than for any mathematical reason. Maynard [36] tackles this brilliantly by introducing a number of new ideas, including ideas from the geometry of numbers, different aspects of sieve theory, and comparisons with a Markov process. We may expect that even in base 3 there should be infinitely many primes with a given digit missing; in base 2, the only digit that might be omitted is 0, and we find the problem of whether there are infinitely many Mersenne primes, which lies beyond reasonable mathematics. We close this discussion by pointing out two other beautiful results on the digits of prime numbers which have elements in common with Maynard’s work: namely, work of Mauduit and Rivat [31] which shows (in particular) that the sum of the decimal digits of primes is equally likely to be odd or even, and work of Bourgain [6] which allows one to specify a small proportion of the binary digits of primes.

**Gaps between primes.** We now turn to a discussion of Maynard’s most spectacular result — the sun amidst small stars — namely, Theorem 1 above on finding many primes in bounded intervals. To describe the recent history of this problem, let us first discuss how primes are spaced typically. The prime number theorem tells us that the  $n$ -th prime  $p_n$  is about  $n \log n$ , so that the average spacing between two consecutive primes,  $p_{n+1} - p_n$ , is about  $\log p_n$ . What is the distribution of the normalized spacings  $(p_{n+1} - p_n)/\log p_n$ ? The Cramér random model for primes would predict that these normalized spacings should behave like a Poisson process, and that for any fixed interval  $[\alpha, \beta] \in \mathbb{R}_{\geq 0}$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \#\left\{n \leq N : \frac{p_{n+1} - p_n}{\log p_n} \in [\alpha, \beta]\right\} = \int_{\alpha}^{\beta} e^{-t} dt = e^{-\alpha} - e^{-\beta}. \quad (4)$$

Gallagher [16] showed that this prediction is also implied by the more refined Hardy–Littlewood conjectures, the key point being that the singular series constants  $\mathfrak{S}(\mathcal{H})$  (see (2)) are approximately 1 (matching the naive Cramér model) on average over  $k$ -element sets  $\mathcal{H}$ .

This conjecture on the normalized spacings between primes is wide open. Indeed if we denote by  $\mathcal{L}$  the set of limit points of  $(p_{n+1} - p_n)/\log p_n$ , then even the qualitative statement that  $\mathcal{L} = [0, \infty]$  (which follows at once from (4)) is currently unknown. By creating long strings of composite numbers, Westzynthius established that  $\mathcal{L}$  contains  $\infty$ , but for a long time no other limit point was known (although Erdős and Ricci had established that  $\mathcal{L}$  has positive Lebesgue measure). Dramatic progress was made in the 2005 with the path-breaking work of Goldston, Pintz, and Yıldırım [17], who showed that for any  $\epsilon > 0$  there are infinitely many  $n$  with  $p_{n+1} - p_n \leq \epsilon \log p_n$ . Thus there are small gaps between primes in comparison to the average, and 0 is now known to be in  $\mathcal{L}$ . Before the work of Goldston, Pintz, and Yıldırım, it was only known that the difference between consecutive primes became smaller than about  $\frac{1}{4}$  of the average spacing, and their work opened the door to later advances including Maynard’s Theorem 1.

Suppose  $h_1, \dots, h_k$  are distinct integers with  $\mathfrak{S}(\{h_1, \dots, h_k\}) > 0$ ; such tuples are called *admissible*, and for example  $\{k!, 2 \cdot k!, \dots, k \cdot k!\}$  is admissible. The Hardy–Littlewood conjecture predicts that there are infinitely many  $n$  with  $n + h_1, \dots, n + h_k$  all being prime. Instead of wanting all  $k$  of these numbers to be prime, what if we only ask for at least two of them to be prime? This would already show that infinitely often there are bounded gaps between consecutive prime numbers. Suppose we could find non-negative weights  $w(n)$  with the property that for large  $x$  and each  $j = 1, \dots, k$ ,

$$\sum_{\substack{x \leq n \leq 2x \\ n+h_j \text{ prime}}} w(n) > \frac{1}{k} \sum_{x \leq n \leq 2x} w(n). \quad (5)$$

Then summing (5) over all  $j = 1, \dots, k$  we would obtain

$$\sum_{x \leq n \leq 2x} \#\{1 \leq j \leq k : n + h_j \text{ prime}\} w(n) > \sum_{x \leq n \leq 2x} w(n), \quad (6)$$

from which it would follow that there must be some  $n$  with at least 2 primes among  $n + h_1, \dots, n + h_k$ . Thinking of the weights as giving a probability measure on  $x \leq n \leq 2x$ , we may

interpret (6) as saying that the expected number of primes among the  $n + h_j$  is greater than 1, so that there must be  $n$  with at least 2 primes in this  $k$ -tuple.

The difficult problem is to construct weights satisfying (5), and natural choices for such weights are suggested by sieve theory, in particular the theory of the Selberg sieve. The standard choice of Selberg sieve weights (which are used to give an upper bound for the number of prime  $k$ -tuples  $n + h_1, \dots, n + h_k$ ) takes the shape

$$w(n) = \left( \sum_{\substack{d|(n+h_1)\cdots(n+h_k) \\ d \leq R}} \mu(d) \left( \frac{\log R/d}{\log R} \right)^k \right)^2.$$

Clearly  $w(n) \geq 0$  always. Expanding out the sum, the right side of (5) (the sum over all  $n \in [x, 2x]$ ) may be evaluated asymptotically so long as  $R^2 \leq x^{1-\epsilon}$ . The left side of (5) is more involved, and relies on understanding the distribution of primes in arithmetic progressions with the modulus of the progression going up to  $R^2$ . The Bombieri–Vinogradov theorem permits such an understanding (at a level comparable to what the Generalized Riemann Hypothesis would give) so long as  $R^2 \leq x^{\frac{1}{2}-\epsilon}$ , so that  $R$  is now constrained to be  $\leq x^{\frac{1}{4}-\epsilon}$ . For this choice of weights, the expected number of primes among the  $n + h_j$  turns out to be about  $(2k/(k+1)) \log R/\log x$ , so that with  $R \leq x^{1/4-\epsilon}$  one only expects to find  $\frac{1}{2}$  a prime in the  $k$ -tuple.

Although the Selberg sieve weights described above had been optimized for upper bounds in the prime  $k$ -tuple problem, Goldston, Pintz, and Yıldırım made the surprising discovery that there are better choices of weights for optimizing the ratio of the sums in (5). They considered weights of the form

$$w(n) = \left( \sum_{\substack{d|(n+h_1)\cdots(n+h_k) \\ d \leq R}} \mu(d) \left( \frac{\log R/d}{\log R} \right)^{k+\ell} \right)^2,$$

for a suitable parameter  $\ell$ , which turns out in the optimal case to be around  $\sqrt{k}$ . With this choice of weights, they found that the expected number of primes among  $n + h_j$  is about twice as large as previously, being  $(4 + O(1/k^{\frac{1}{2}})) \log R/\log x$ . With  $R = x^{\frac{1}{4}-\epsilon}$ , this barely fails to give the desired relation (5), and thus barely falls short of proving bounded gaps between primes. By considering an additional possible prime value  $n + h$  for  $1 \leq h \leq \epsilon \log x$ , Goldston, Pintz, Yıldırım were able to deduce from this argument that there are infinitely many  $n$  with  $p_{n+1} - p_n \leq \epsilon \log n$ . For a more detailed discussion of these ideas see [47].

If one could take  $R$  to be  $x^{\frac{1}{4}+\delta}$  for any  $\delta > 0$ , then the argument of Goldston, Pintz, and Yıldırım would give bounded gaps between primes. To take such a value for  $R$ , one would need to understand the distribution of primes up to  $x$  in arithmetic progressions, when the modulus of the progression is as large as  $x^{\frac{1}{2}+2\delta}$ . The Elliott–Halberstam conjectures predict that such results should hold (on average) when the modulus is as large as  $x^{1-\epsilon}$ . Partial progress towards such extensions of the Bombieri–Vinogradov theorem was made by Fouvry and Iwaniec [12], and Bombieri, Friedlander, and Iwaniec [5], but these results did not apply immediately to the problem of showing bounded gaps between primes. In April 2013, Yitang Zhang [49] made a spectacular breakthrough by establishing a version of the



Bombieri–Vinogradov theorem in an extended range which was sufficient for the method of Goldston, Pintz, and Yıldırım. Zhang established that if  $k > 3.5 \times 10^6$  then for any admissible  $k$ -tuple  $h_1, \dots, h_k$  there are infinitely many  $n$  with at least two of the  $n + h_j$  being prime. This implied that infinitely often the gaps between consecutive primes is less than 70 million. Refinements of Zhang’s work on the equidistribution of primes in arithmetic progressions were made by the Polymath project [46], and still further qualitative and quantitative refinements of such results may be found in the recent papers of Maynard [37–39].

Zhang’s work established the case  $m = 2$  of Theorem 1. However, even if one could take the largest possible range for  $R$ , namely  $R = x^{\frac{1}{2}-\epsilon}$  (which would be permitted by the Elliott–Halberstam conjecture), the Goldston–Pintz–Yıldırım weights would only yield that the expected number of primes in an admissible  $k$ -tuple is  $\geq 2 - \epsilon$ . In other words, even under the Elliott–Halberstam conjecture one would fall short of establishing the existence of three primes in bounded intervals.

The proof of Theorem 1 is based on a different choice of the weights  $w(n)$ , discovered just months after Zhang’s work by Maynard (who announced the results in a memorable talk at Oberwolfach in October 2013) and independently by Tao (in unpublished work). The Maynard–Tao weights are a multi-dimensional extension of the weights considered earlier, and take (roughly speaking) the shape

$$w(n) = \left( \sum_{\substack{d_1, \dots, d_k \\ d_i | n + h_i \\ \prod d_i \leq R}} \prod_{i=1}^k \mu(d_i) F\left(\frac{\log d_1}{\log R}, \dots, \frac{\log d_k}{\log R}\right) \right)^2,$$

for suitable smooth functions  $F : [0, 1]^k \rightarrow \mathbb{R}$ . Astonishingly it turns out that for an appropriate choice for  $F$ , the expected number of primes in the tuple  $n + h_1, \dots, n + h_k$  (recall (6) above) is  $\geq c \log k \frac{\log R}{\log x}$ , for a positive constant  $c$ ; in fact  $c$  may be taken close to 1 if  $k$  is large enough. The key point is that this expected number of primes in  $k$ -tuples tends to infinity with  $k$ , and in fact we only need  $R$  to grow like any power of  $x$  for the method to succeed, so that Bombieri–Vinogradov which permits  $R = x^{\frac{1}{4}-\epsilon}$  is already sufficient! Thus the following more precise version of Theorem 1 holds, which may be viewed as a partial result towards the Hardy–Littlewood prime  $k$ -tuples conjecture.

**Theorem 3** (Maynard [32]). *Let  $m \geq 2$  be a natural number. Let  $k$  be sufficiently large in terms of  $m$ , and let  $\mathcal{H} = \{h_1, \dots, h_k\}$  be any set of  $k$  integers with  $\mathfrak{S}(\mathcal{H}) > 0$ . Then there exist infinitely many  $n$  such that the  $k$ -tuple  $n + h_1, \dots, n + h_k$  contains at least  $m$  primes.*

Maynard showed that  $k$  may be taken smaller than  $Cm^2 e^{4m}$  for a suitable constant  $C$ , and further refinements of this (incorporating also the work of Zhang) have been made in the work of Baker and Irving [3] who showed that  $k$  may be taken as  $Ce^{3.815m}$ . Of special interest is the case  $m = 2$  where the Polymath project [45] optimized these arguments to establish that any admissible 50-tuple contains 2 primes infinitely often. In particular, they showed that  $p_{n+1} - p_n \leq 246$  infinitely often, and conditional on the Elliott–Halberstam conjecture that infinitely often there are at least two primes in the triple  $n, n + 2, n + 6$ . Let us mention one other uniform variant of these results: Maynard [33] shows, for instance, that there are at least

$cX \exp(-\sqrt{\log X})$  values of  $x \in [X, 2X]$  such that the interval  $[x, x + \log X]$  contains at least  $c \log \log X$  primes (here  $c$  is a positive constant). For detailed expositions on these results of Zhang, Maynard, and Tao, see [18, 29].

The Maynard–Tao weights offer a flexible new method to study many problems on primes and related sequences, and have found a number of applications. We describe two other results using these weights, both still concerned with spacings between consecutive primes. We referred earlier to the result of Westzynthius on large gaps between consecutive primes, which showed that  $\infty$  lies in the set  $\mathcal{L}$  of limit points of the normalized spacings  $(p_{n+1} - p_n)/\log p_n$ . This was quantified in the 1930’s by Erdős and Rankin who showed that, for a positive constant  $C$

$$\max_{p_n \leq X} (p_{n+1} - p_n) \geq C \log X \frac{(\log \log X) \log \log \log \log X}{(\log \log \log X)^2}. \quad (7)$$

The random model would suggest that the maximal gap between primes up to  $X$  should be about  $(\log X)^2$ . This is known as Cramér’s conjecture, and while this is very delicate, it is widely believed that the maximal gap is no more than  $(\log X)^{2+\epsilon}$ , although even this is far beyond Landau’s unattainable problem of the existence of a prime between consecutive squares. Erdős drew attention to the problem of finding larger gaps between consecutive primes, offering \$ 10 000 for a bound that would replace  $C$  in (7) with a function tending to  $\infty$  with  $X$ . For more than 75 years, this problem resisted attack, with only improvements of the constant  $C$  being known. Then, by a remarkable coincidence, in 2014 *two* different techniques emerged, both establishing (7) with  $C$  replaced by a function tending to infinity with  $X$ . One approach, by Ford, Green, Konyagin, and Tao [11], built upon the work of Green–Tao on arithmetic progressions in the primes, while the other approach, by Maynard [34], found a way to adapt the Maynard–Tao sieve weights. The second approach was better suited for quantifying the large gaps that are produced, and, joining forces, Ford, Green, Konyagin, Maynard, and Tao [10] established that for some constant  $C > 0$

$$\max_{p_n \leq X} (p_{n+1} - p_n) \geq C \log X \frac{(\log \log X) \log \log \log \log X}{\log \log \log X}, \quad (8)$$

improving the bound in (7) by a factor of  $\log \log \log X$ .

The results on small gaps and large gaps between consecutive primes show that 0 and  $\infty$  lie in the set  $\mathcal{L}$  of limit points of the normalized prime spacings. No other explicit numbers are known to lie in  $\mathcal{L}$ , although we expect  $\mathcal{L}$  to include all non-negative real numbers. Following Zhang’s breakthrough, Pintz [42] showed that  $\mathcal{L}$  contains an interval  $[0, c]$  for some  $c > 0$ , which however is ineffective and cannot be computed explicitly. Using the Maynard–Tao sieve weights, Banks, Freiberg, and Maynard [4] established the following beautiful result: If  $\beta_1 \leq \beta_2 \leq \dots \leq \beta_9$  are any nine real numbers, then at least one of their differences  $\beta_j - \beta_i$  (with  $i < j$ ) must be an element of  $\mathcal{L}$ . Their result has been refined by Pintz [43], and Merikoski [41], and Merikoski shows that the same result holds if we start with just four real numbers  $\beta_1 \leq \beta_2 \leq \beta_3 \leq \beta_4$ . Moreover, Merikoski has also shown that for any  $T > 0$ , the set  $\mathcal{L} \cap [0, T]$  has measure at least  $T/3$ .

**The Duffin–Schaeffer conjecture.** So far we have focussed entirely on Maynard’s work concerned with prime numbers. In a very different direction, Maynard in joint work with Koukoulopoulos [28], resolved one of the central problems in the metric theory of Diophantine approximation, known as the Duffin–Schaeffer conjecture.

Diophantine approximation is concerned with finding rational approximations  $a/q$  to a given irrational number  $\alpha$ , with an emphasis on making  $|\alpha - a/q|$  small in terms of  $q$ . The most basic result is Dirichlet’s theorem that for every irrational number  $\alpha$ , there are infinitely many rational approximations  $a/q$ , with  $a \in \mathbb{Z}$ ,  $q \in \mathbb{N}$  and  $(a, q) = 1$  (so that the fraction is in reduced form) such that  $|\alpha - a/q| \leq 1/q^2$ . For quadratic irrationals (like  $\sqrt{2}$  or the golden ratio), Dirichlet’s theorem is essentially the best possible, and for every such  $\alpha$  there exists a positive constant  $C(\alpha)$  such that  $|\alpha - a/q| \geq C(\alpha)/q^2$  for any rational approximation  $a/q$ . A celebrated result of Roth establishes that for any algebraic irrational  $\alpha$  and any  $\epsilon > 0$  one has  $|\alpha - a/q| \geq C(\alpha, \epsilon)/q^{2+\epsilon}$ , for a suitable positive constant  $C(\alpha, \epsilon)$ . For particular interesting transcendental numbers, such as  $\pi$ , it remains an outstanding open problem to determine how well they can be approximated by rational numbers.

Metric Diophantine approximation is concerned with such approximation problems that hold for *almost all* irrational numbers  $\alpha$ , with *almost all* interpreted in the sense of Lebesgue measure. Since the problem of approximating  $\alpha$  by rationals is identical to that of approximating  $\alpha + 1$ , we may restrict attention to irrational numbers  $\alpha \in [0, 1)$ . The most basic problem is the following: suppose  $\psi : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$  is a given function, what can be said about the measure of  $\alpha \in [0, 1)$  for which there exist infinitely many rational numbers  $a/q$  in reduced form (that is,  $(a, q) = 1$ ) with  $|\alpha - a/q| \leq \psi(q)$ . For instance, Dirichlet’s theorem tells us that if  $\psi(q) = 1/q^2$ , then all irrational  $\alpha \in [0, 1)$  admit infinitely many such rational approximations.

Let  $\mathcal{A}_q = \mathcal{A}_q(\psi)$  denote the set of  $\alpha \in [0, 1)$  for which there exists some reduced fraction  $a/q$  with  $|\alpha - a/q| \leq \psi(q)$ , and let  $\mathcal{A}$  denote the set of  $\alpha \in [0, 1)$  lying in infinitely many of the sets  $\mathcal{A}_q$ . Thus

$$\mathcal{A} = \bigcap_{Q=1}^{\infty} \tilde{\mathcal{A}}(Q), \quad \text{with} \quad \tilde{\mathcal{A}}(Q) = \bigcup_{q=Q}^{\infty} \mathcal{A}_q.$$

Now the measure of  $\mathcal{A}_q$  is  $\leq 2\phi(q)\psi(q)$ , since there are  $\phi(q)$  possible choices for the numerator  $a$ , and if  $\psi(q) \leq 1/(2q)$  so that the intervals for different  $a$  do not overlap then equality holds here. If  $\sum_{q=1}^{\infty} \phi(q)\psi(q)$  converges, then the measure of  $\tilde{\mathcal{A}}(Q)$  is bounded by  $2 \sum_{q=Q}^{\infty} \phi(q)\psi(q)$ , which is the tail of a convergent series and thus tends to 0 as  $Q \rightarrow \infty$ . It follows that  $\mathcal{A}$  has measure 0. This argument is identical to the easy part of the Borel–Cantelli Lemma.

In 1941, Duffin and Schaeffer made the remarkable conjecture that in the complementary case when  $\sum_{q=1}^{\infty} \phi(q)\psi(q)$  diverges, the measure of  $\mathcal{A}$  is 1. Since then the Duffin–Schaeffer conjecture has remained one of the central motivating questions in the theory of metric Diophantine approximations. A number of partial results towards this conjecture were established: for example, a beautiful result of Gallagher [15] showed that the measure of the set  $\mathcal{A}(\psi)$  is always either 0 or 1, work of Erdős [9] and Vaaler [48] established

the conjecture when  $\psi(q)$  is  $O(1/q^2)$  for all  $q$ , higher dimensional analogues of the conjecture were proved by Pollington and Vaughan [44], and weaker versions of the conjecture with extra divergence conditions were established in [1, 22, 23]. But the full problem resisted until the recent work of Koukoulopoulos and Maynard [28]:

**Theorem 4** (Koukoulopoulos and Maynard [28]). *Let  $\psi : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$  be such that  $\sum_{q=1}^{\infty} \phi(q)\psi(q)$  diverges. Then the set of  $\alpha \in [0, 1)$  that have infinitely many rational approximations  $|\alpha - a/q| \leq \psi(q)$  with  $(a, q) = 1$  has Lebesgue measure 1. In other words, the Duffin–Schaeffer conjecture holds.*

We refer to Koukoulopoulos’s talk at this ICM [27] for a more detailed exposition of this result, and the ideas behind its proof.

We have given an overview of some of Maynard’s most spectacular achievements in analytic number theory. Maynard’s work is characterized by ingenious but simple ideas, which are carried very far with his powerful technical ability. As impressive as his work so far has been, it may only mark a beginning.

### Funding

This work was partially supported by grants from the National Science Foundation, and a Simons Investigator Award from the Simons Foundation.

### References

- [1] C. Aistleitner, T. Lachmann, M. Munsch, N. Technau, and A. Zafeiropoulos, The Duffin-Schaeffer conjecture with extra divergence. *Adv. Math.* **356** (2019), 106808, 11
- [2] R. C. Baker, G. Harman, and J. Pintz, The difference between consecutive primes. II. *Proc. London Math. Soc. (3)* **83** (2001), no. 3, 532–562
- [3] R. C. Baker and A. J. Irving, Bounded intervals containing many primes. *Math. Z.* **286** (2017), no. 3-4, 821–841
- [4] W. D. Banks, T. Freiberg, and J. Maynard, On limit points of the sequence of normalized prime gaps. *Proc. Lond. Math. Soc. (3)* **113** (2016), no. 4, 515–539
- [5] E. Bombieri, J. B. Friedlander, and H. Iwaniec, Primes in arithmetic progressions to large moduli. *Acta Math.* **156** (1986), no. 3-4, 203–251
- [6] J. Bourgain, Prescribing the binary digits of primes, II. *Israel J. Math.* **206** (2015), no. 1, 165–182
- [7] C. Dartyge and C. Mauduit, Nombres presque premiers dont l’écriture en base  $r$  ne comporte pas certains chiffres. *J. Number Theory* **81** (2000), no. 2, 270–291
- [8] C. Dartyge and C. Mauduit, Ensembles de densité nulle contenant des entiers possédant au plus deux facteurs premiers. *J. Number Theory* **91** (2001), no. 2, 230–255

- [9] P. Erdős, On the distribution of the convergents of almost all real numbers. *J. Number Theory* **2** (1970), 425–441
- [10] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao, Long gaps between primes. *J. Amer. Math. Soc.* **31** (2018), no. 1, 65–105
- [11] K. Ford, B. Green, S. Konyagin, and T. Tao, Large gaps between consecutive prime numbers. *Ann. of Math. (2)* **183** (2016), no. 3, 935–974
- [12] E. Fouvry and H. Iwaniec, On a theorem of Bombieri-Vinogradov type. *Mathematika* **27** (1980), no. 2, 135–152 (1981)
- [13] J. Friedlander and H. Iwaniec, The polynomial  $X^2 + Y^4$  captures its primes. *Ann. of Math. (2)* **148** (1998), no. 3, 945–1040
- [14] J. Friedlander and H. Iwaniec, *Opera de cribro*. American Mathematical Society Colloquium Publications 57, American Mathematical Society, Providence, RI, 2010
- [15] P. Gallagher, Approximation by reduced fractions. *J. Math. Soc. Japan* **13** (1961), 342–345
- [16] P. X. Gallagher, On the distribution of primes in short intervals. *Mathematika* **23** (1976), no. 1, 4–9
- [17] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, Primes in tuples. I. *Ann. of Math. (2)* **170** (2009), no. 2, 819–862
- [18] A. Granville, Primes in intervals of bounded length. *Bull. Amer. Math. Soc. (N.S.)* **52** (2015), no. 2, 171–222
- [19] B. Green and T. Tao, Linear equations in primes. *Ann. of Math. (2)* **171** (2010), no. 3, 1753–1850
- [20] B. Green and T. Tao, The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)* **175** (2012), no. 2, 541–566
- [21] B. Green, T. Tao, and T. Ziegler, An inverse theorem for the Gowers  $U^{s+1}[N]$ -norm. *Ann. of Math. (2)* **176** (2012), no. 2, 1231–1372
- [22] G. Harman, *Metric number theory*. London Mathematical Society Monographs. New Series 18, The Clarendon Press, Oxford University Press, New York, 1998
- [23] A. K. Haynes, A. D. Pollington, and S. L. Velani, The Duffin-Schaeffer conjecture with extra divergence. *Math. Ann.* **353** (2012), no. 2, 259–273
- [24] D. R. Heath-Brown, Primes represented by  $x^3 + 2y^3$ . *Acta Math.* **186** (2001), no. 1, 1–84
- [25] D. R. Heath-Brown and X. Li, Prime values of  $a^2 + p^4$ . *Invent. Math.* **208** (2017), no. 2, 441–499
- [26] H. A. Helfgott, The ternary Goldbach problem. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. II*, pp. 391–418, Kyung Moon Sa, Seoul, 2014
- [27] D. Koukoulopoulos, Rational approximations of irrational numbers. 2021, URL <https://arxiv.org/abs/2109.11003>
- [28] D. Koukoulopoulos and J. Maynard, On the Duffin-Schaeffer conjecture. *Ann. of Math. (2)* **192** (2020), no. 1, 251–307

- [29] E. Kowalski, Gaps between prime numbers and primes in arithmetic progressions [after Y. Zhang and J. Maynard]. *Astérisque* (2015), no. 367-368, Exp. No. 1084, ix, 327–366
- [30] K. Matomäki, J. Maynard, and X. Shao, Vinogradov’s theorem with almost equal summands. *Proc. Lond. Math. Soc. (3)* **115** (2017), no. 2, 323–347
- [31] C. Mauduit and J. Rivat, Sur un problème de Gelfond: la somme des chiffres des nombres premiers. *Ann. of Math. (2)* **171** (2010), no. 3, 1591–1646
- [32] J. Maynard, Small gaps between primes. *Ann. of Math. (2)* **181** (2015), no. 1, 383–413
- [33] J. Maynard, Dense clusters of primes in subsets. *Compos. Math.* **152** (2016), no. 7, 1517–1554
- [34] J. Maynard, Large gaps between primes. *Ann. of Math. (2)* **183** (2016), no. 3, 915–933
- [35] J. Maynard, Digits of primes. In *European Congress of Mathematics*, pp. 641–661, Eur. Math. Soc., Zürich, 2018
- [36] J. Maynard, Primes with restricted digits. *Invent. Math.* **217** (2019), no. 1, 127–218
- [37] J. Maynard, Primes in arithmetic progressions to large moduli I: Fixed residue classes. 2020, URL <https://arxiv.org/abs/2006.06572>
- [38] J. Maynard, Primes in arithmetic progressions to large moduli II: Well-factorable estimates. 2020, URL <https://arxiv.org/abs/2006.07088>
- [39] J. Maynard, Primes in arithmetic progressions to large moduli III: Uniform residue classes. 2020, URL <https://arxiv.org/abs/2006.08250>
- [40] J. Maynard, Primes represented by incomplete norm forms. *Forum Math. Pi* **8** (2020), e3, 128
- [41] J. Merikoski, Limit points of normalized prime gaps. *J. Lond. Math. Soc. (2)* **102** (2020), no. 1, 99–124
- [42] J. Pintz, Polignac numbers, conjectures of Erdős on gaps between primes, arithmetic progressions in primes, and the bounded gap conjecture. In *From arithmetic to zeta-functions*, pp. 367–384, Springer, [Cham], 2016
- [43] J. Pintz, A note on the distribution of normalized prime gaps. *Acta Arith.* **184** (2018), no. 4, 413–418
- [44] A. D. Pollington and R. C. Vaughan, The  $k$ -dimensional Duffin and Schaeffer conjecture. *Mathematika* **37** (1990), no. 2, 190–200
- [45] D. H. J. Polymath, New equidistribution estimates of Zhang type. *Algebra Number Theory* **8** (2014), no. 9, 2067–2199
- [46] D. H. J. Polymath, Variants of the Selberg sieve, and bounded intervals containing many primes. *Res. Math. Sci.* **1** (2014), Art. 12, 83
- [47] K. Soundararajan, Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım. *Bull. Amer. Math. Soc. (N.S.)* **44** (2007), no. 1, 1–18
- [48] J. D. Vaaler, On the metric theory of Diophantine approximation. *Pacific J. Math.* **76** (1978), no. 2, 527–539

[49] Y. Zhang, Bounded gaps between primes. *Ann. of Math. (2)* **179** (2014), no. 3, 1121–1174

**Kannan Soundararajan**

Department of Mathematics, Stanford University, Stanford CA 94305,  
[ksound@stanford.edu](mailto:ksound@stanford.edu)