A Dynamical Proof of the Prime Number Theorem

Redmond Mcnamara

Hardy-Ramanujan Journal 44 (2021), 160-185 submitted 22/05/2020, accepted 01/02/2021, revised 26/04/2021

A Dynamical Proof of the Prime Number Theorem

Redmond McNamara

Abstract. We present a new, elementary, dynamical proof of the prime number theorem.

Keywords. prime number theorem, entropy decrement argument
2010 Mathematics Subject Classification. 11A41, 11N05, 37A45

1. Introduction

The prime number theorem states that

$$\pi(N) = (1 + o_{N\to\infty}(1))\frac{N}{\log N},$$

where $\pi(N)$ denotes the number of primes of size at most N. In some sense, the result was first publicly conjectured by Legendre in 1798 who suggested that

$$\pi(N) = \frac{N}{A\log N + B + o_{N\to\infty}(1)},$$

for some constants A and B. Legendre specifically conjectured $A = 1$ and $B = -1.08366$. Gauss conjectured the same formula and stated he was not sure what the constant B might turn out to be. Gauss' conjecture was based on millions of painstaking calculations first obtained in 1792 and 1793 which were never published but nonetheless predate Legendre's work on the subject. It is worth noting that later in his 1849 letter to Encke Gauss conjectured that $\pi(N) \approx \mathrm{Li}(N)$, which in particular implies the correct values for A and B. The first major breakthrough on the problem was due to Chebyshev who showed that

$$c + o_{N\to\infty}(1) \leq \frac{\pi(N)\log N}{N} \leq C + o_{N\to\infty}(1)$$

for some explicit constants c and C with $c > 0$. There is a long history of improvements to these explicit constants for which we refer to Goldstein [Gol73] and Goldfeld [Gol04]. The prime number theorem was important motivation for Riemann's seminal work on the zeta function.

The first proofs of the prime number theorem were given independently by Hadamard and de la Vallée Poussin in 1896. The key step in their proof is a difficult argument showing that the Riemann zeta function does not have a zero on the line $\mathrm{Re}(z) = 1$. Their proof was later substantially simplified by many mathematicians. In 1930, Wiener found a "Fourier analytic" proof of the prime number theorem. In 1949, Erdös [Erd49] and Selberg [Sel50] discovered an elementary proof of the prime number theorem, where here elementary is used in the technical sense that the proof involves no complex analysis and does not necessarily mean that the proof is easy reading. The bitter battle over credit for this result is the subject of an informative note by Goldfeld [Gol04]. Other proofs are due to Daboussi [Dab89] and Hildebrand [Hil86]. In a blog post from 2014, Tao proves the prime number theorem using the theory of Banach algebras [Taob]. A published version of this theorem can be found in a book by Einsiedler and Ward [EW17]. In an unpublished book from 2014, Granville and

We thank episciences.org for providing open access hosting of the electronic journal *Hardy-Ramanujan Journal*

Soundarajan prove the prime number theorem using pretentious methods (see, for instance, [GHS19]). A note by Zagier [Zag97] from 1997 contains perhaps the quickest proof of the prime number theorem using a tauberian argument in the spirit of the Erdös-Selberg proof combined with complex analysis in the form of Cauchy's theorem. Zagier attributes this proof to Newman.

The goal of this note is to present a new proof of the prime number theorem. Florian Richter and I discovered similar proofs concurrently and independently. His proof can be found in [Ric]. Terence Tao wrote up a version of this argument on his blog following personal communication from the author which can be found in [Taoa].

The proof proceeds as follows. To prove the prime number theorem, it suffices to prove that

$$\frac{1}{N}\sum_{n\leq N}\Lambda(n) = 1 + o(1),$$

where $\Lambda(n)$ is the von Mangoldt function which is $\log p$ if n is a power of a prime p and 0 otherwise. The reader may think of Λ as the normalized indicator function of the primes. The von Mangoldt function is related to the Möbius function via the formula

$$\Lambda = \mu * \log,$$

where the Möbius function $\mu(n)$ is 0 if n has a repeated factor, -1 if n has an odd number of distinct prime factors, $+1$ if n has an even number of distinct prime factors. This formula, sometimes called the Möbius inversion formula, encodes the fundamental theorem of arithmetic. Thus, there is a dictionary between properties of the von Mangoldt function Λ and the Möbius function μ. Landau observed that cancellation in the Möbius function is equivalent to the prime number theorem i.e. the prime number theorem is equivalent to the statement

$$\frac{1}{N}\sum_{n\leq N}\mu(n) = o_{N\to\infty}(1).$$

This is what we actually try to prove.

The next observation is that, if one wants to compute a sum, it suffices to sample only a small number of terms. Typically (for instance for an i.i.d. randomly chosen sequence) the average value

$$\frac{1}{N}\sum_{n\leq N}a(n)$$

is approximately the same as the average over only the even terms

$$\approx\frac{2}{N}\sum_{n\leq N}a(n)\mathbb{1}_{2|n}.$$

However, for certain sequences, like $a = (-1, +1, -1, +1, \ldots)$, the averages do not agree. Still for this sequence, if we instead sample every third point or every fifth point or every p^{th} point for any other prime then the averages are approximately equal. It turns out, this is a rather general phenomenon: for any sequence, for most primes p, the average of the sequence is the same as the average along only those numbers divisible by p.

Applying this to the Möbius function, for each N, for most primes p

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx \frac{1}{N}\sum_{n\leq N}\mu(n)p\mathbb{1}_{p|n}.$$

For the purposes of this introduction, we will "cheat" and pretend that this equation is true for any prime p. By changing variables

$$\frac{1}{N}\sum_{n\le N}\mu(n)p\mathbb{1}_{p|n} = \frac{p}{N}\sum_{n\le N/p}\mu(pn).$$

But $\mu(pn) = -\mu(n)$ for most numbers n since μ is multiplicative. Combining the last two equations gives

$$\frac{1}{N}\sum_{n\le N}\mu(n) \approx -\frac{p}{N}\sum_{n\le N/p}\mu(n).$$

The plan is to use this identity three times. Suppose we can find primes p_1, p_2 and p such that $\frac{p_1p_2}{p} \approx 1$. Then by applying the previous identity

$$\frac{1}{N}\sum_{n\le N}\mu(n) \approx -\frac{p}{N}\sum_{n\le N/p}\mu(n)$$

and also

$$\begin{aligned}\frac{1}{N}\sum_{n\le N}\mu(n) &\approx -\frac{p_1}{N}\sum_{n\le N/p_1}\mu(n)\\ &\approx +\frac{p_1p_2}{N}\sum_{n\le N/p_1p_2}\mu(n).\end{aligned}$$

But since $\frac{p_1p_2}{p} \approx 1$, we know that

$$\frac{p_1p_2}{N}\sum_{n\le N/p_1p_2}\mu(n) \approx \frac{p}{N}\sum_{n\le N/p}\mu(n).$$

Putting everything together we conclude that

$$\frac{1}{N}\sum_{n\le N}\mu(n) \approx -\frac{1}{N}\sum_{n\le N}\mu(n)$$

which implies

$$\frac{1}{N}\sum_{n\le N}\mu(n) \approx 0.$$

This implies the prime number theorem.

Thus, the main difficulty in the proof is finding primes p, p_1 and p_2 lying outside some exceptional set for which $\frac{p_1p_2}{p} \approx 1$. We give a quick sketch of the argument. The Selberg symmetry formula (Theorem 2.5) roughly tells us that, even if we do not know how many primes there are at a certain scale (say in the interval from x to $x(1+\varepsilon)$) and we do not know how many semiprimes (products of two primes) there are at that scale, the weighted sum of the number of primes and semiprimes is as we would expect. In particular, if there are no semiprimes between x and $x(1+\varepsilon)$ there are twice as many primes as one would expect (meaning $2\cdot\varepsilon\frac{x}{\log x}$ many primes). Let x be a large number. If there are both primes and semiprimes between x and $x(1+\varepsilon)$ then we can find p, p_1 and p_2 such that $\frac{p_1p_2}{p} \approx 1+O(\varepsilon)$ and we are done. Thus, assume that there are either only primes or only semiprimes in the interval $[x, x(1+\varepsilon)]$. For the sake of our exposition, we will assume there are only primes between x and $x(1+\varepsilon)$. By the Selberg symmetry formula, there are twice as many primes in this interval as expected. Now if there is a semiprime p_1p_2 in the interval $[x(1+\varepsilon), x(1+\varepsilon)^2]$ then

picking any prime p in the interval $[x, x(1+\varepsilon)]$ we conclude that there exists p, p_1 and p_2 such that $\frac{p_1 p_2}{p} \approx 1 + O(\varepsilon)$. Thus, either we win (and the prime number theorem is true) or there are again twice as many primes in the interval $[x(1+\varepsilon), x(1+\varepsilon)^2]$ as one would expect. Running this argument again shows that there are again only primes and no semiprimes in the interval $[x(1+\varepsilon)^2, x(1+\varepsilon)^3]$. Iterating this argument using the connectedness of the interval, we find large intervals $[x, 100x]$ where there are twice as many primes as predicted by the prime number theorem. But this contradicts Chebyshev's theorem: Chebyshev's theorem gives a lower bound on the number of primes, which in turn gives a lower bound on the number of semiprimes; alternately, we remark that one could use Erdös's version of Chebyshev's theorem that the number of primes less than x is at most $\log 4 \frac{x}{\log x}$ and because $\log 4 < 2$ this gives a contradiction. This completes the proof.

1.A. A comment on notation

Throughout this paper, we will use asymptotic notation. Since number theory, dynamics and analysis sometime use different conventions, we take a moment here to fix notation. We will write

$$x = O(y)$$

to mean that there exists a constant C such that

$$|x| \leq Cy.$$

When we adorn these symbols with subscripts, the subscripts specify which variables the constants are allowed to depend on. Thus

$$x = O_{A,B}(y)$$

means that there exists a constant C which is allowed to depend on A and B such that

$$|x| \leq Cy.$$

We write

$$x = y + O(z)$$

to mean that

$$x - y = O(z).$$

We also adopt little o notation:

$$x = o_{n\to\infty}(y)$$

means that

$$\lim_{n\to\infty} \frac{x}{y} = 0.$$

Occasionally, when the variable with respect to which the limit is being taken is clear from context, we may simply write

$$x = o(y).$$

As before, we write

$$x = y + o_{n\to\infty}(z)$$

to mean

$$x - y = o_{n\to\infty}(z).$$

If the expression x depends on more than one variables, say n, m and k, we may use subscripts to make explicit that the rate of convergence implicit in the little o notation is allowed to depend on more variables. Thus,

$$x = o_{n\to\infty,m,k}(y)$$

means that $\frac{x}{y}$ tends to zero with n at a rate which may depend on m and k.

2. Proof of the prime number theorem

From number theory, we will use Mertens' Theorem, in particular the version which states,

$$\sum_{p\leq x} \frac{1}{p} = \log\log x + M + O\left(\frac{1}{\log x}\right)$$

for some constant M; we will also use Chebyshev's Theorem, the Selberg Symmetry Formula, Landau's formulation of the prime number theorem (i.e. that the prime number theorem is equivalent to $\sum_{n\leq N} \mu(n) = o_{N\to\infty}(N)$) and a slightly modified version of the Turán-Kubilius inequality which we will prove using the following Bombieri-Halász-Montgomery inequality.

Proposition 2.1. (Bombieri-Halász-Montgomery inequality [Bom71]) *Let w_i be a sequence of nonnegative real numbers. Let u and v_i be vectors in a Hilbert space. Then*

$$\sum_{i=1}^{n} w_i |\langle u, v_i\rangle|^2 \leq ||u||^2 \cdot \left(\sup_i \sum_{j=1}^{n} w_j |\langle v_i, v_j\rangle|\right).$$

Proof. By duality, there exists c_i such that

$$\sum_{i=1}^{n} w_i |c_i|^2 = 1$$

and

$$\sum_{i=1}^{n} w_i |\langle u, v_i\rangle|^2 = \left(\sum_{i=1}^{n} w_i c_i \langle u, v_i\rangle\right)^2$$

and therefore by conjugate bilinearity of the inner product

$$= \left\langle u, \sum_{i=1}^{n} w_i \overline{c}_i v_i \right\rangle^2.$$

By Cauchy-Schwarz, this is at most

$$\leq ||u||^2 \left|\left| \sum_{i=1}^{n} w_i \bar{c}_i v_i \right|\right|^2.$$

By the pythagorean theorem this is given by

$$= ||u||^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j c_i \bar{c}_j \langle v_i, v_j \rangle.$$

The geometric mean is dominated by the arithmetic mean.

$$\leq ||u||^2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \frac{1}{2}(|c_i| + |c_j|) |\langle v_i, v_j \rangle|.$$

By symmetry this is

$$= ||u||^2 \sum_{i=1}^{n} w_i |c_i|^2 \sum_{j=1}^{n} w_j |\langle v_i, v_j \rangle|.$$

Because everything is nonnegative, we may replace the inner term with a supremum

$$\leq ||u||^2 \sum_{i=1}^{n} w_i |c_i|^2 \sup_k \sum_{j=1}^{n} w_j |\langle v_k, v_j \rangle|.$$

Using that $\sum w_i |c_i|^2 = 1$ completes the proof.

The next proposition applies the previous proposition in order to show that, for any bounded sequence, the average of the sequence is the same as the average over the p^{th} terms in the sequence for most prime p.

Proposition 2.2. (Turán-Kubilius [Kub64]) *Let S denote a set of primes less than some natural number P. Let N be a natural number which is at least P^3. Let f be a 1-bounded function from $\mathbb{N}$ to $\mathbb{C}$. Then*

$$\sum_{p \in S} \frac{1}{p} \left| \frac{1}{N} \sum_{n \leq N} f(n)(1 - p\mathbb{1}_{p|n}) \right|^2 = O(1).$$

Proof. We will apply Proposition 2.1: our Hilbert space is L^2 on the space of function on the integers $\{1, \ldots, N\}$ equipped with normalized counting measure; set $w_p = \frac{1}{p}$; set $v_p = (n \mapsto 1 - p\mathbb{1}_{p|n})$ and $u = f$; thus, by Proposition 2.1

$$\sum_{p \in S} \frac{1}{p} \left| \frac{1}{N} \sum_{n \leq N} f(n)(1 - p\mathbb{1}_{p|n}) \right|^2$$

$$\leq \frac{1}{N} \sum_{n \leq N} |f(n)|^2 \cdot \sup_{p \in S} \sum_{q \in S} \frac{1}{q} \left| \frac{1}{N} \sum_{n \leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) \right|.$$

Since f is 1-bounded, we may bound the L^2 norm of f by 1. Thus,

$$\leq \sup_{p\in S} \sum_{q\in S} \frac{1}{q} \left| \frac{1}{N} \sum_{n\leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) \right|. \tag{2.1}$$

For primes p and q,

$$\frac{1}{N} \sum_{n\leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n})$$

can be expanded into a signed sum of four terms

$$\frac{1}{N} \sum_{n\leq N} 1 - p\mathbb{1}_{p|n} - q\mathbb{1}_{q|n} + pq\mathbb{1}_{p|n}\mathbb{1}_{q|n}.$$

When $p \neq q$, we claim that each term is $1 + O\left(\frac{P^2}{N}\right)$. The trickiest term is the last term

$$\frac{1}{N} \sum_{n\leq N} pq\mathbb{1}_{p|n}\mathbb{1}_{q|n}.$$

When $p \neq q$, we have that

$$\mathbb{1}_{p|n}\mathbb{1}_{q|n} = \mathbb{1}_{pq|n}.$$

Of course, for any natural number m,

$$\#\text{ of } n \leq N \text{ such that } m \text{ divides } n \ = \frac{N}{m} + O(1),$$

where the $O(1)$ term comes from the fact that m need not perfectly divide N. Thus,

$$\frac{1}{N} \sum_{n\leq N} pq\mathbb{1}_{p|n}\mathbb{1}_{q|n} = pq\left(\frac{1}{pq} + O\left(\frac{1}{N}\right)\right),$$

which is $1 + O\left(\frac{P^2}{N}\right)$ as claimed. A similar argument handles the three other terms. Altogether, we conclude that

$$\frac{1}{N} \sum_{n\leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) = O\left(\frac{P^2}{N}\right),$$

when $p \neq q$. Inserting this bound into 2.1 and remembering that there are at most P terms in the sum over q in S, we find

$$\begin{aligned}
&\sum_{p\in S} \frac{1}{p} \left| \frac{1}{N} \sum_{n\leq N} f(n)(1 - p\mathbb{1}_{p|n}) \right|^2 \\
&\leq \sup_{p\in S} \sum_{q\in S} \frac{1}{q} \left| \frac{1}{N} \sum_{n\leq N} (1 - p\mathbb{1}_{p|n})(1 - q\mathbb{1}_{q|n}) \right| \\
&\leq \sup_{p\in S} \frac{1}{p} \left| \frac{1}{N} \sum_{n\leq N} (1 - p\mathbb{1}_{p|n})(1 - p\mathbb{1}_{p|n}) \right| + O\left(\frac{P^3}{N}\right)
\end{aligned}$$

Expanding out the product, the main term is

$$\sup_{p\in S}\frac{1}{p}\left|\frac{1}{N}\sum_{n\leq N}p^2\mathbb{1}_{p|n}\right|.$$

By the same trick as before, we may replace the average of $\mathbb{1}_{p|n}$ by $\frac{1}{p}$ plus a small error dominated by the main term. Cancelling factors of p as appropriate, we are left with

$$=\sup_{p\in S}\frac{1}{p}\left|\frac{1}{N}\sum_{n\leq N}p^2\frac{1}{p}\right|=O(1).$$

Of course, all the smaller terms can be bounded by the triangle inequality. This completes the proof.

Note that,

$$\sum_{p\in S}\frac{1}{p}\frac{1}{N}\sum_{n\leq N}1=\sum_{p\in S}\frac{1}{p}.$$

For instance, if S is the set of all primes less than P, Euler proved that

$$\sum_{p\leq P}\frac{1}{p}\to\infty$$

as P tends to infinity. In fact, Mertens' theorem states that this sum is approximately $\log\log P$. Thus, Proposition 2.2 represents a real improvement over the trivial bound. Therefore, for S, P, N and f as in the statement of Proposition 2.2

$$\left|\frac{1}{N}\sum_{n\leq N}f(n)(1-p\mathbb{1}_{p|n})\right|^2$$

is small for "most" primes. This shows that most primes are "good" in the sense that

$$\frac{1}{N}\sum_{n\leq N}f(n)\approx\frac{1}{N}\sum_{n\leq N}f(n)p\mathbb{1}_{p|n}$$

This notion is captured in the following definition.

Definition 2.3. *Let ε be a positive real number, let P be a natural number which is sufficiently large depending on ε and let N be a natural number sufficiently large depending on P. Denote by $\ell(N)$ the quantity*

$$\ell(N)=\sum_{n\leq N}\frac{1}{n}.$$

Denote by $S(N)$ the set of primes $p\leq P$ such that

$$\frac{1}{N}\left|\sum_{n\leq N}\mu(n)-\sum_{n\leq N}\mu(n)p\mathbb{1}_{p|n}\right|\geq\varepsilon.$$

Then we say a prime p is good if

$$\frac{1}{\ell(N)}\sum_{n\leq N}\frac{1}{n}\mathbb{1}_{p\in S(n)}\leq\varepsilon.$$

Otherwise, we say p is bad.

From Proposition 2.2, we obtain the following corollary.

Corollary 2.4. *Let ε be a positive real number, let P be a natural number which is sufficiently large depending on ε and let N be a natural number sufficiently large depending on P. Then the set of bad primes is small in the sense that*

$$\sum_{p \text{ bad } \leq P} \frac{1}{p} = O(\varepsilon^{-3}).$$

Proof. By Proposition 2.2, for each n sufficiently large,

$$\sum_{p \leq P} \frac{1}{p} \mathbb{1}_{p \notin S(n)} = O(\varepsilon^{-2}).$$

Summing in n gives,

$$\sum_{p \leq P} \frac{1}{p} \frac{1}{\ell(N)} \sum_{n \leq N} \frac{1}{n} \mathbb{1}_{p \notin S(n)} = O(\varepsilon^{-2}) + o_{N \to \infty, P}(1).$$

We remark that for N sufficiently large depending on P, this second error term may be absorbed into the first term. By definition, the set of bad primes is the set of primes such that

$$\frac{1}{\ell(N)} \sum_{n \leq N} \frac{1}{n} \mathbb{1}_{p \notin S(n)} \geq \varepsilon.$$

But then by Chebyshev's inequality (i.e. not his theorem on counting primes),

$$\sum_{p \text{ bad } \leq P} \frac{1}{p} = O(\varepsilon^{-3}).$$

as desired.

Next, we turn to the Selberg symmetry formula. To state Selberg's symmetry formula, we need to introduce the following function. Let $\Lambda_2 = \log \cdot \Lambda + \Lambda * \Lambda$ i.e.

$$\Lambda_2(n) = \log(n)\Lambda(n) + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right),$$

where the von Mangoldt function $\Lambda(n)$ when $\log p$ is n is a power of a prime p and 0 otherwise. Thus, we remark that Λ_2 is supported on prime powers and products of two prime powers. It is not too hard to show that Λ_2 is "mostly" supported on primes and semiprimes. Recall that the prime number theorem is the statement that

$$\frac{1}{N} \sum_{n \leq N} \Lambda(n) = 1 + o_{N \to \infty}(1)$$

and thus

$$\frac{1}{N} \sum_{n \leq N} \Lambda(n) \log n = \log(N)(1 + o_{N \to \infty}(1)).$$

We are now ready to state the Selberg symmetry formula.

Theorem 2.5. (Selberg symmetry formula) *The average of the second von Mangoldt function defined above is*

$$\frac{1}{N} \sum_{n \leq N} \Lambda_2(n) = 2 \log N(1 + o_{N \to \infty}(1)).$$

We will refer the reader to, for instance, [Taob] section 1 for the proof. The next proposition says that, at each scale, there are either many primes or many semiprimes.

Proposition 2.6. *Let $\varepsilon > 0$ be a sufficiently small number. Suppose that k_0 is sufficiently large depending on ε and let I_k denote the interval $[(1+\varepsilon)^k, (1+\varepsilon)^{k+1}]$. Then for every $k \geq k_0$,*

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} \geq \frac{1}{k}.$$

Proof. This follows from the Selberg symmetry formula (Theorem 2.5): after all, by the Selberg symmetry formula, for k_0 sufficiently large, for all $k \geq k_0$,

$$\frac{1}{(1+\varepsilon)^k} \sum_{n \leq (1+\varepsilon)^k} \Lambda_2(n) = 2 \log(1+\varepsilon)^k (1 + O(\varepsilon^2)).$$

The same holds for k replaced by $k+1$.

$$\frac{1}{(1+\varepsilon)^{k+1}} \sum_{n \leq (1+\varepsilon)^{k+1}} \Lambda_2(n) = 2 \log(1+\varepsilon)^{k+1} (1 + O(\varepsilon^2)).$$

Taking differences, and using that $k \log(1+\varepsilon) = (k+1)\log(1+\varepsilon)(1+O(\varepsilon^2))$, for $k \geq k_0$ sufficiently large, we find that

$$\frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{n \in I_k} \Lambda_2(n) = 2 \log(1+\varepsilon)^k (1 + O(\varepsilon)). \tag{2.2}$$

We aim to show that prime powers do not contribute very much to this sum. Notice that, if a prime power contributes to the sum, then the corresponding prime must be at most the square root of $(1+\varepsilon)^{k+1}$ and there is at most one power of any prime in the interval I_k (because $\varepsilon < 1$). Also, notice that $\Lambda_2(p^a) \leq 2\Lambda(p^a)\log p^a$. Thus, we bound

$$\begin{aligned}
\frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{\substack{n = p^a, a > 1 \\ n \in I_k}} \Lambda(n) \log n &= \frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{\substack{n = p^a, a > 1 \\ n \in I_k}} \log p \log p^a \\
&\leq \frac{1}{\varepsilon(1+\varepsilon)^k} \sum_{p \leq (1+\varepsilon)^{(k+1)/2}} \log p \log (1+\varepsilon)^{k+1}.
\end{aligned}$$

Now the number of primes less than $(1+\varepsilon)^{(k+1)/2}$ is certainly less than $(1+\varepsilon)^{(k+1)/2}$, so

$$\begin{aligned}
&\leq \frac{1}{\varepsilon(1+\varepsilon)^k} (1+\varepsilon)^{(k+1)/2} \log(1+\varepsilon)^{(k+1)/2} \log(1+\varepsilon)^{k+1}. \\
&= o_{k \to \infty, \varepsilon}(1).
\end{aligned}$$

For instance, by choosing k_0 large depending on ε, we can make this quantity

$$= O(\varepsilon)$$

Similarly for products of a natural number m and a prime power,

$$\begin{aligned}\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{n=p^a m, a>1\\ n\in I_k}}\Lambda(p)\Lambda(m) \leq& \frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{n=p^a m, a>1\\ n\in I_k}}\log p\log m\\ \leq& \frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{p\leq(1+\varepsilon)^{(k+1)/2}}\log p\sum_{1<a\leq\log_p(1+\varepsilon)^k}\sum_{mp^a\in I_k}\Lambda(m).\end{aligned}$$

Now the inner most sum is bounded by Chebyshev's inequality. We simplify slightly using the factor of $\frac{1}{\varepsilon(1+\varepsilon)^k}$ out front.

$$\begin{aligned}\leq& C\sum_{p\leq(1+\varepsilon)^{(k+1)/2}}\log p\sum_{1<a\leq\log_p(1+\varepsilon)^k}\frac{1}{p^a}.\\ \leq& C\sum_{p\leq(1+\varepsilon)^{(k+1)/2}}\frac{\log p}{p^2}.\\ =& O(1).\end{aligned}$$

Finally, we claim that when one of the prime factors of a semiprime is less than $\exp(\varepsilon^3 k)$ then that semiprime does not contribute very much to the sum. Indeed,

$$\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{p_1p_2\in I_k\\ p_1\leq\exp(\varepsilon^3 k)}}\Lambda(p_1)\Lambda(p_2)=\frac{\varepsilon}{(1+\varepsilon)^k}\sum_{\substack{p_1p_2\in I_k\\ p_1\leq\exp(\varepsilon^3 k)}}\log p_1\log p_2.$$

Now we use that p_1 is at most $\exp(\varepsilon^3 k)$ and p_2 is at most $(1+\varepsilon)^{k+1}$.

$$\leq\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{p_1p_2\in I_k\\ p_1\leq\exp(\varepsilon^3 k)}}\varepsilon^3 k\cdot\log(1+\varepsilon)^{k+1}.$$

Summing over scales,

$$\leq\frac{1}{\varepsilon(1+\varepsilon)^k}\varepsilon^3 k\cdot\log(1+\varepsilon)^{k+1}\sum_{m\leq k\varepsilon^3}\sum_{\substack{m\leq\log p_1\leq m-1\\ p_1p_2\in I_k}}1.$$

The number of terms in the inner sum is can be estimated using Chebyshev's theorem. The outersum has roughly $k\varepsilon^3$ many terms. Thus, for some constant C,

$$\leq C\frac{1}{\varepsilon(1+\varepsilon)^k}\cdot\varepsilon^3 k\log(1+\varepsilon)^{k+1}\cdot\varepsilon^3 k\cdot\frac{\exp(\varepsilon^3 k)}{\varepsilon^3 k}\frac{(1+\varepsilon)^{k+1}}{\log(1+\varepsilon)^{k+1}}$$

Simplifying, this is

$$\begin{aligned}=& O(\varepsilon^2 k)\\ =& O(\varepsilon\cdot\log(1+\varepsilon)^k).\end{aligned}$$

Altogether, we find that we can restrict 2.2 to primes and semiprimes where neither factor is too small.

$$\frac{1}{\varepsilon(1+\varepsilon)^k}\sum_{\substack{n\in I_k\\ n=p\text{ or }n=p_1p_2\\ p_i\geq\exp(\varepsilon^3 k)}}\Lambda_2(n)=2\log(1+\varepsilon)^k(1+O(\varepsilon)).$$

For any two numbers n and m in I_k, $\frac{1}{n} = \frac{1}{m} \cdot (1 + O(\varepsilon))$, so

$$\varepsilon^{-1} \sum_{\substack{n \in I_k \\ n=p \text{ or } n=p_1p_2 \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{\Lambda_2(n)}{n} = 2\log(1+\varepsilon)^k(1+O(\varepsilon)).$$

By the pigeonhole principle, either

$$\varepsilon^{-1} \sum_{p \in I_k} \frac{\Lambda_2(p)}{p} \geq \log(1+\varepsilon)^k(1+O(\varepsilon))$$

or

$$\varepsilon^{-1} \sum_{\substack{p_1p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{\Lambda_2(p_1p_2)}{p_1p_2} \geq \log(1+\varepsilon)^k(1+O(\varepsilon)).$$

In the first case, moving the ε and $\log p \approx \log(1+\varepsilon)^k$ terms to the other side

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{\varepsilon}{k\log(1+\varepsilon)} \cdot (1+O(\varepsilon)).$$

Taylor expanding the logarithm gives

$$\geq \frac{1}{k} \cdot (1+O(\varepsilon)),$$

as desired. In the second case,

$$\varepsilon^{-1} \sum_{\substack{p_1p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1p_2} k^2 \log^2(1+\varepsilon) \geq \log(1+\varepsilon)^k(1+O(\varepsilon)).$$

Rearranging terms gives

$$\sum_{\substack{p_1p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1p_2} \geq \frac{\varepsilon}{k\log(1+\varepsilon)}(1+O(\varepsilon)).$$

Taylor expanding the logarithm again completes the proof.

Next, we show that we can actually find two nearby scales where both inequalities from Proposition 2.6 hold. The key idea is to use the connectedness of the interval.

Proposition 2.7. *Let $\varepsilon > 0$ be a number sufficiently small. Suppose that k_0 is sufficiently large depending on ε and let I_k denote the interval $(1+\varepsilon)^k$ to $(1+\varepsilon)^{k+1}$. Then there exists k and k' such that $|k-k'| \leq 1$ with k and k' in $[k_0, \varepsilon^{-2}+k_0]$ and such that*

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{2k}$$

and

$$\sum_{\substack{p_1p_2 \in I_{k'} \\ p_i \geq \exp(\varepsilon^3 k')}} \frac{1}{p_1p_2} \geq \frac{1}{2k'}$$

Proof. Suppose not. Then by Proposition 2.6, for each k in $[k_0, \varepsilon^{-2} + k_0]$ either

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{2k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} \geq \frac{1}{2k}.$$

If both hold for some k, then by choosing $k = k'$, we could conclude that Proposition 2.7 holds. Thus, we will assume that exactly one of

$$\sum_{p \in I_k} \frac{1}{p} \geq \frac{1}{2k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} \geq \frac{1}{2k}$$

hold for any choice of k. Whichever holds for k_0 must also hold for $k_0 + 1$ since otherwise we may choose $k = k_0$ and $k' = k_0 + 1$. Inductively, we may assume that for every k in $[k_0, \varepsilon^{-2} + k_0]$ either

$$\sum_{p \in I_k} \frac{1}{p} < \frac{1}{2k}$$

or

$$\sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} < \frac{1}{2k}.$$

Summing in k, we eventually obtain a contradiction with Mertens' theorem: either

$$\sum_{(1+\varepsilon)^{k_0} \leq p \leq (1+\varepsilon)^{k_0+\varepsilon^{-2}}} \frac{1}{p} < \frac{1}{10} \cdot \left(\log(k_0 + \varepsilon^{-2}) - \log k_0 + O\left(\frac{1}{k_0}\right)\right) \tag{2.3}$$

or

$$\sum_{k \in [k_0, \varepsilon^{-2}+k_0]} \sum_{\substack{p_1 p_2 \in I_k \\ p_i \geq \exp(\varepsilon^3 k)}} \frac{1}{p_1 p_2} < \frac{1}{2} \cdot \left(\log(k_0 + \varepsilon^{-2}) - \log k_0 + O\left(\frac{1}{k_0}\right)\right). \tag{2.4}$$

We remark that a Taylor expansion could simplify

$$\log(k_0 + \varepsilon^{-2}) - \log k_0 + O\left(\frac{1}{k_0}\right) = O\left(\frac{1}{\varepsilon^2 k_0}\right).$$

Note that Mertens' theorem implies that

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + M + O\left(\frac{1}{\log x}\right),$$

for some constant M. Taking differences,

$$\begin{aligned} \sum_{(1+\varepsilon)^{k_0} \leq p \leq (1+\varepsilon)^{k_0+\varepsilon^{-2}}} \frac{1}{p} &= \log\log(1+\varepsilon)^{k_0+\varepsilon^{-2}} - \log\log(1+\varepsilon)^{k_0} + O\left(\frac{1}{k_0 \log(1+\varepsilon)}\right) \\ &= \log(k_0 + \varepsilon^{-2}) - \log(k_0) + O\left(\frac{1}{k_0 \log(1+\varepsilon)}\right). \end{aligned}$$

But 2.3 says that the sum on the left is 2 times smaller than that which gives a contradiction.

In the next proposition, we show that this implies there are nearby primes and semiprimes which are good.

Proposition 2.8. *Let $\varepsilon > 0$. Let P be a natural number which is sufficiently large depending on ε. Let N be a natural number which is sufficiently large depending on P. Then there exists p_1, p_2 and p such that*

$$\frac{p_1 p_2}{p} = 1 + O(\varepsilon)$$

with p_1, p_2 and p good in the sense of Definition 2.3 meaning p_1, p_2 and p are not in $S(n)$ for "most" $n \leq N$ (see Definition 2.3 for details). Furthermore, we can require that p_1, p_2 and p are all greater than $\frac{1}{\varepsilon}$.

Proof. By Proposition 2.7, it suffices to show that, for some k_0 sufficiently large depending on ε with the property that $(1+\varepsilon)^{k_0+\varepsilon^{-2}} \leq P$, we have

$$\sum_{\substack{p \in [(1+\varepsilon)^{k_0}, (1+\varepsilon)^{\varepsilon^{-2}+k_0}] \\ p \text{ bad}}} \frac{1}{p} \leq \frac{1}{10k_0} \tag{2.5}$$

and that

$$\sum_{\substack{p_1 p_2 \in [(1+\varepsilon)_0^k, (1+\varepsilon)^{\varepsilon^{-2}+k_0}] \\ p_1 \text{ bad} \\ p_1^{\varepsilon^3} \leq p_2 \leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1 p_2} \leq \frac{1}{10k_0}. \tag{2.6}$$

After all, once we have shown this, we can argue as follows: by Proposition 2.7 there exists an interval of the form k and k' in $[k_0, k_0 + \varepsilon^{-2}]$ with $|k - k'| \leq 1$ for which

$$\sum_{p \in [(1+\varepsilon)^k, (1+\varepsilon)^{k+1}]} \frac{1}{p} > \frac{1}{k}$$

and

$$\sum_{\substack{p_1 p_2 \in I_{k'} \\ p_i \geq \exp(\varepsilon^3 k')}} \frac{1}{p_1 p_2} \geq \frac{1}{k'},$$

where $I_k = [(1+\varepsilon)^k, (1+\varepsilon)^{k+1}]$ and similarly for $I_{k'}$. By 2.5

$$\sum_{\substack{p \in [(1+\varepsilon)^k, (1+\varepsilon)^{k+1}] \\ p \text{ good}}} \frac{1}{p} > 0$$

and by 2.6

$$\sum_{\substack{p_1 p_2 \in I_{k'} \\ p_i \geq \exp(\varepsilon^3 k') \\ p_1, p_2 \text{ good}}} \frac{1}{p_1 p_2} > 0.$$

Now any good p in I_k and any good $p_1 p_2$ in $I_{k'}$ suffices to prove the result.

Now, for the sake of contradiction, suppose first that

$$\sum_{\substack{p \in [(1+\varepsilon)_0^k, (1+\varepsilon)^{\varepsilon^{-2}+k_0}] \\ p \text{ bad}}} \frac{1}{p} \geq \frac{1}{10k_0}.$$

Summing in $k_0 \leq \log\log P$, for P large enough we get that

$$\sum_{\substack{p\leq \log N\\ p \text{ bad}}} \frac{1}{p} \geq \frac{1}{20}\log\log\log P$$

which contradicts Corollary 2.4. Second, suppose that

$$\sum_{\substack{p_1p_2\in[(1+\varepsilon)_0^k,(1+\varepsilon)^{\varepsilon^{-2}+k_0}]\\ p_1 \text{ bad}\\ p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1p_2} \geq \frac{1}{10k_0}.$$

Summing in $k_0 \leq \log\log P$ gives, for P large enough

$$\sum_{\substack{p_1p_2\leq \log N\\ p_1 \text{ bad}\\ p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1p_2} \geq \frac{1}{20}\log\log\log P.$$

For each p_1, by Mertens' theorem,

$$\sum_{p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}} \frac{1}{p_2} \leq -10\log\varepsilon.$$

By Corollary 2.4, this implies

$$\sum_{\substack{p_1p_2\leq \log N\\ p_1 \text{ bad}\\ p_1^{\varepsilon^3}\leq p_2\leq p_1^{\varepsilon^{-3}}}} \frac{1}{p_1p_2} = O\left(\varepsilon^{-3}|\log\varepsilon|\right)$$

which yields a contradiction since for P large enough, $\frac{1}{20}\log\log\log P \gg \varepsilon^{-3}|\log\varepsilon|$.

Finally, we show this implies the prime number theorem.

Theorem 2.9. *The prime number theorem holds, i.e.*

$$\frac{1}{N}\sum_{n\leq N}\Lambda(n) = 1 + o_{N\to\infty}(1)$$

Proof. Let ε be a positive real number, let P be a natural number which is sufficiently large depending on ε and let N be a natural number sufficiently large depending on P. By Proposition 2.8, there exist primes p_1, p_2 and p all good and greater than $\frac{1}{\varepsilon}$ such that

$$\frac{p_1p_2}{p} = 1 + O(\varepsilon).$$

By definition of a good prime,

$$\frac{1}{M}\left|\sum_{n\leq M}\mu(n) - \sum_{n\leq M}\mu(n)p\mathbb{1}_{p|n}\right| \geq \varepsilon,$$

for at most a small set of M (exactly how small will be spelled out shortly). In particular, let $S(M)$ denote the set of primes such that

$$\frac{1}{M}\left|\sum_{n\leq M}\mu(n)-\sum_{n\leq M}\mu(n)p\mathbb{1}_{p|n}\right|\geq\varepsilon.$$

Then by definition of a good prime,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\mathbb{1}_{p_1\in S(M)}\mathbb{1}_{p_2\in S(M)}\mathbb{1}_{p\in S(M)}=O(\varepsilon).$$

Thus, we may conclude that

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\frac{1}{M}\left|\sum_{n\leq M}\mu(n)-\sum_{n\leq M}\mu(n)p\mathbb{1}_{p|n}\right|=O(\varepsilon).$$

Since $\mu(np)=-\mu(n)$ for most n (including all but those $O(\frac{1}{p})=O(\varepsilon)$ fraction of n which are not divisible by p), we conclude that

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)+\frac{p}{M}\sum_{n\leq M/p}\mu(n)\right|=O(\varepsilon).$$

Similarly, since p_1 is good,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)+\frac{p_1}{M}\sum_{n\leq M/p_1}\mu(n)\right|=O(\varepsilon).$$

By change of variables,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{p_1}{M}\sum_{n\leq M/p_1}\mu(n)+\frac{p_1p_2}{M}\sum_{n\leq M/p_1p_2}\mu(n)\right|=O(\varepsilon)+O\left(\frac{\log p_1}{\log N}\right).$$

By the triangle inequality and since N is much larger than p_1,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{p}{M}\sum_{n\leq M/p}\mu(n)+\frac{p_1p_2}{M}\sum_{n\leq M/p_1p_2}\mu(n)\right|=O(\varepsilon).$$

But since $\frac{p_1p_2}{p}=1+O(\varepsilon)$,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{p}{M}\sum_{n\leq M/p}\mu(n)\right|=O(\varepsilon).$$

and therefore, again using that p is good,

$$\frac{1}{\ell(N)}\sum_{M\leq N}\frac{1}{M}\left|\frac{1}{M}\sum_{n\leq M}\mu(n)\right|=O(\varepsilon).$$

This is an averaged version on the equation we want. We want that

$$\left|\frac{1}{N}\sum_{n\leq N}\mu(n)\right|=O(\varepsilon),$$

for all N sufficiently large. Thus, we just need to prove

Lemma 2.10. *Let $\varepsilon > 0$, let N be sufficiently large depending on ε and suppose that*

$$\frac{1}{\ell(N)} \sum_{M \leq N} \frac{1}{M} \left| \frac{1}{M} \sum_{n \leq M} \mu(n) \right| = O(\varepsilon).$$

Then

$$\left| \frac{1}{N} \sum_{n \leq N} \mu(n) \right| = O(\varepsilon),$$

To prove this we use the identity

$$\mu \cdot \log = -\mu * \Lambda.$$

Summing both sides up to N gives

$$\sum_{n \leq N} \mu(n) \log n = - \sum_{n \leq N} \sum_{d | n} \mu\left(\frac{n}{d}\right) \Lambda(d).$$

Now by switching the order of summation

$$= - \sum_{d \leq N} \Lambda(d) \left(\sum_{n \leq N/d} \mu(n) \right).$$

If it were not for the factor of $\Lambda(d)$, this would be exactly what we want. Each $\sum_{n \leq M} \mu(n)$ for an integer M occurs in this sum the number of times that $\left\lfloor \frac{N}{d} \right\rfloor = M$ where $\lfloor \cdot \rfloor$ denotes the floor which is proportional to $\frac{N}{M^2}$. The factor of $\Lambda(d)$ can be removed using the Brun-Titchmarsh inequality as follows. First, we break up the sum into different scales

$$= - \sum_{a \in (1+\varepsilon)^{\mathbb{N}}} \sum_{\substack{d \leq N \\ a \leq d < (1+\varepsilon)a}} \Lambda(d) \left(\sum_{n \leq N/d} \mu(n) \right).$$

For all d between a and $(1+\varepsilon)a$, the sums $\sum_{n \leq N/d} \mu(n)$ all give roughly the same value. Therefore

$$= - \sum_{a \in (1+\varepsilon)^{\mathbb{N}}} \sum_{\substack{d \leq N \\ a \leq d < (1+\varepsilon)a}} \Lambda(d) \left(\sum_{n \leq N/a} \mu(n) \right)$$

$$+ O\left(\sum_{\substack{a \in (1+\varepsilon)^{\mathbb{N}} \\ a \leq N}} \sum_{a \leq d < (1+\varepsilon)a} \Lambda(d) \sum_{\frac{N}{a(1+\varepsilon)} \leq n \leq \frac{N}{a}} 1 \right)$$

First, we focus on the error term.

$$
\begin{aligned}
&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\sum_{\frac{N}{a(1+\varepsilon)}\leq n\leq\frac{N}{a}}1\right)\\
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\left(\frac{N}{a}-\frac{N}{a(1+\varepsilon)}\right)\right)\\
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\sum_{a\leq d<(1+\varepsilon)a}\Lambda(d)\frac{N\varepsilon}{a(1+\varepsilon)}\right)
\end{aligned}
$$

By the Brun-Titchmarsh inequality

$$
\begin{aligned}
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\varepsilon a\frac{N\varepsilon}{a(1+\varepsilon)}\right)\\
=&O\left(\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}N\varepsilon^2\right)\\
=&O\left(N\varepsilon^2\log_{1+\varepsilon}N\right)\\
=&O\left(\varepsilon N\log N\right)
\end{aligned}
$$

where the last step involves Taylor expanding $\log(1+\varepsilon)$ near $\varepsilon=0$. Next, we turn our attention to the main term. We begin by pulling out the sum over $\mu(n)$ which no longer depends on d.

$$
\begin{aligned}
&-\sum_{a\in(1+\varepsilon)^{\mathbb{N}}}\sum_{\substack{d\leq N\\ a\leq d<(1+\varepsilon)a}}\Lambda(d)\left(\sum_{n\leq N/a}\mu(n)\right)\\
=&-\sum_{a\in(1+\varepsilon)^{\mathbb{N}}}\left(\sum_{n\leq N/a}\mu(n)\right)\left(\sum_{\substack{d\leq N\\ a\leq d<(1+\varepsilon)a}}\Lambda(d)\right).
\end{aligned}
$$

By the Brun-Titchmarsh inequality, this is bounded in absolute value by

$$
\leq\sum_{\substack{a\in(1+\varepsilon)^{\mathbb{N}}\\ a\leq N}}\left|\sum_{n\leq N/a}\mu(n)\right|\cdot(10\varepsilon a).
$$

Earlier, we replaced a sum indexed by $n\leq N/d$ by a sum indexed by $n\leq N/a$, showing these two sums were close up to an error of size $O(\varepsilon N\log N)$. Undoing this process, we find

$$
=10\sum_{d\leq N}\left|\sum_{n\leq N/d}\mu(n)\right|+O(\varepsilon N\log N).
$$

Now we let $M = \frac{N}{d}$. The number of values of d such that $\lfloor \frac{N}{d} \rfloor$ is the number of values of d such that $M \leq \frac{N}{d} < M+1$ and therefore $\frac{N}{M+1} < d \leq \frac{N}{M}$. The number of such d's is bounded by $\frac{N}{M} - \frac{N}{M+1} = \frac{N}{M(M+1)}$. Thus

$$\leq 10 \sum_{M \leq N} \frac{N}{M^2} \left| \sum_{n \leq M} \mu(n) \right| + O(\varepsilon N \log N).$$

But we already showed that this sum is bounded by

$$= O(\varepsilon N \ell(N))$$
$$= O(\varepsilon N \log N).$$

Thus,

$$\sum_{n \leq N} \mu(n) \log(n) = O(\varepsilon N \log N).$$

Since $\log n = \log N(1 + O(\varepsilon))$ for n between $\varepsilon \frac{N}{\log N}$ and N and ε sufficiently small we conclude that

$$\sum_{n \leq N} \mu(n) = O(\varepsilon N).$$

But this classically implies the prime number theorem.

3. In what ways is this a dynamical proof?

To begin the argument, we showed that for all N, for most p i.e. all p outside a bad set where

$$\sum_{p \text{ bad}} \frac{1}{p} \leq C_\varepsilon$$

we have that

$$\sum_{n \leq N} \mu(n) = \sum_{n \leq N} \mu(n) p \mathbb{1}_{p|n} + O(\varepsilon).$$

We did this using an L^2 orthogonality argument (Propositions 2.1 and 2.2). Alternately, we can argue using a variant of Tao's entropy decrement argument (the first version of this argument appeared in [Tao16]; a different version of the entropy decrement argument appeared in [TT18] and [TT19]; the version presented here is somewhat different from what appeared in those papers). Let $\mathbf{n}$ be a random integer less than N. Let $\mathbf{x}_i = \mu(\mathbf{n}+i)$ and let $\mathbf{y}_p = \mathbf{n} \bmod p$. In probability and dynamics, a stochastic process is a sequence of random variables $(\ldots, \xi_{-2}, \xi_{-1}, \xi_0, \xi_1, \xi_2, \ldots)$ such that

$$\mathbb{P}((\xi_1, \ldots \xi_k) \in A) = \mathbb{P}((\xi_{1+m}, \ldots \xi_{k+m}) \in A)$$

for any set A and for any m. In our setting $(\ldots, \mathbf{x}_{-2}, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots)$ is approximately stationary in the sense that

$$\mathbb{P}((\mathbf{x}_1, \ldots \mathbf{x}_k) \in A) \approx \mathbb{P}((\mathbf{x}_{1+m}, \ldots \mathbf{x}_{k+m}) \in A)$$

where the two terms differ by some small error which is $o_{N \to \infty, m}(1)$. A stationary process is the same as a random variable in a measure preserving system where ξ_{i+1} is the transformation applied to ξ_i. A key invariant of a stationary process is thus the Kolmogorov-Sinai entropy:

$$h(\xi) = \lim_{n \to \infty} \frac{1}{n} H(\xi_1, \ldots, \xi_n)$$

where

$$H(\xi_1, \dots, \xi_n)$$

is the Shannon entropy of $(\xi_1, \dots, \xi_n)$. This limit exists because

$$\frac{1}{n} H(\xi_1, \dots, \xi_n) = \frac{1}{n} \sum_{i \leq n} H(\xi_i | \xi_1 \dots, \xi_{i-1})$$

by the chain rule for entropy, which is equal to

$$= \frac{1}{n} \sum_{i \leq n} H(\xi_0 | \xi_{-1} \dots, \xi_{-i+1})$$

by stationarity. This is a Caesaró average of a decreasing sequence which is therefore decreasing. Since entropy is nonnegative, we can conclude that the limit exists. In our case, because $(\dots, \mathbf{x}_{-1}, \mathbf{x}_0, \mathbf{x}_1, \dots)$ is almost stationary, we can conclude that

$$\frac{1}{n} H(\mathbf{x}_1, \dots, \mathbf{x}_n)$$

is almost decreasing in the sense that, for $m > n$,

$$\frac{1}{m} H(\mathbf{x}_1, \dots, \mathbf{x}_m) \leq \frac{1}{n} H(\mathbf{x}_1, \dots, \mathbf{x}_n) + o_{N \to \infty, n}(1).$$

The same is true for the relative entropy

$$\frac{1}{n} H(\mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{y}_{p_1}, \dots, \mathbf{y}_{p_k})$$

for any fixed set of primes $p_1, \dots, p_k$.

We define the mutual information between two random variables $\mathbf{x}$ and $\mathbf{y}$ as

$$I(\mathbf{x}; \mathbf{y}) = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y})$$

and more generally the conditional mutual information

$$I(\mathbf{x}; \mathbf{y}|\mathbf{z}) = H(\mathbf{x}|\mathbf{z}) - H(\mathbf{x}|\mathbf{y}, \mathbf{z}).$$

We assume for the rest of the explanation that all random variables take only finitely many values. Mutual information measures how close two random variables are to independent. Two random variables $\mathbf{x}$ and $\mathbf{y}$ are independent if and only if

$$I(\mathbf{x}; \mathbf{y}) = 0.$$

Intuitively, we think of $\mathbf{x}$ and $\mathbf{y}$ as close to independent if the mutual information is small. The crux of the entropy decrement argument is that we can find primes p such that $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ is close to independent of $\mathbf{y}_p$. The argument is as follows. Let $p_1 < p_2 < \dots < p_k$ be a sequence of primes. Consider the relative entropy

$$\begin{aligned} &\frac{1}{p_k} H(\mathbf{x}_1, \dots, \mathbf{x}_{p_k} | \mathbf{y}_{p_1}, \dots, \mathbf{y}_{p_k}) \\ =&\frac{1}{p_k} H(\mathbf{x}_1, \dots, \mathbf{x}_{p_k} | \mathbf{y}_{p_1}, \dots, \mathbf{y}_{p_{k-1}}) - \frac{1}{p_k} I(\mathbf{x}_1, \dots, \mathbf{x}_{p_k}; \mathbf{y}_{p_k} | \mathbf{y}_{p_1}, \dots, \mathbf{y}_{p_{k-1}}) \end{aligned}$$

and because the relative entropy is almost decreasing

$$=\frac{1}{p_{k-1}}H(\mathbf{x}_1,\dots,\mathbf{x}_{p_{k-1}}|\mathbf{y}_{p_1},\dots,\mathbf{y}_{p_{k-1}})-\frac{1}{p_k}I(\mathbf{x}_1,\dots,\mathbf{x}_{p_k};\mathbf{y}_{p_k}|\mathbf{y}_{p_1},\dots,\mathbf{y}_{p_{k-1}})+o(1).$$

Inductively, we find

$$\leq H(\mathbf{x}_1)-\sum_{j\leq k}\frac{1}{p_j}I(\mathbf{x}_1,\dots,\mathbf{x}_{p_j};\mathbf{y}_{p_j}|\mathbf{y}_{p_1},\dots,\mathbf{y}_{p_{j-1}})+o(1)$$

We conclude that the set of bad primes p_j for which

$$I(\mathbf{x}_1,\dots,\mathbf{x}_{p_j};\mathbf{y}_{p_j}|\mathbf{y}_{p_1},\dots,\mathbf{y}_{p_{j-1}})\geq\varepsilon$$

satisfies

$$\sum_{p_j\text{ bad}}\frac{1}{p_j}\leq\varepsilon^{-1}H(\mathbf{x}_1)+o(1)<\infty.$$

Thus, for most primes,

$$I(\mathbf{x}_1,\dots,\mathbf{x}_{p_j};\mathbf{y}_{p_j}|\mathbf{y}_{p_1},\dots,\mathbf{y}_{p_{j-1}})<\varepsilon.$$

In a slight abuse of terminology, we say such primes are good. Although this definition is apparently different from Definition 2.3, we will show that this notion of good meaning small mutual information essentially implies the "random sampling" version defined in Definition 2.3.

Intuitively, if p is good then $\mathbf{x}_1,\dots,\mathbf{x}_p$ and $\mathbf{y}_p$ are nearly independent. This is formalized by Pinsker's inequality. Pinsker's inequality states that

$$d_{TV}(\mathbf{x},\mathbf{y})\leq D(\mathbf{x}||\mathbf{y})^{1/2}$$

where d_{TV} is the total variation distance and D is the Kullback-Leibler divergence. For our purposes, the important thing about the Kullback-Liebler divergence is that if $\mathbf{y}'$ is a random variable with the same distribution as $\mathbf{y}$ which is independent of $\mathbf{x}$ then

$$D((\mathbf{x},\mathbf{y})||(\mathbf{x},\mathbf{y}'))=I(\mathbf{x};\mathbf{y}).$$

Therefore, we conclude that

$$d_{TV}((\mathbf{x},\mathbf{y}),(\mathbf{x},\mathbf{y}'))\leq I(\mathbf{x};\mathbf{y})^{1/2}.$$

Similarly, there is a relative version

$$d_{TV}((\mathbf{x},\mathbf{y},\mathbf{z}),(\mathbf{x},\mathbf{y}',\mathbf{z}))\leq I(\mathbf{x};\mathbf{y}|\mathbf{z})^{1/2},$$

where now $\mathbf{y}'$ has the same distribution as $\mathbf{y}$ but is relatively independent of $\mathbf{x}$ over $\mathbf{z}$ meaning that

$$\mathbb{P}(\mathbf{x}\in A,\mathbf{y}\in B|\mathbf{z}=c)=\mathbb{P}(\mathbf{x}\in A|\mathbf{z}=c)\mathbb{P}(\mathbf{y}\in B|\mathbf{z}=c).$$

Thus, for bounded function F,

$$\mathbb{E}F(\mathbf{x},\mathbf{y},\mathbf{z})=\mathbb{E}F(\mathbf{x},\mathbf{y}',\mathbf{z})+O(I(\mathbf{x};\mathbf{y})^{1/2}),$$

where again $\mathbf{y}'$ is relatively independent of $\mathbf{x}$ over $\mathbf{z}$ and $\mathbb{E}$ denotes the expectation. In our case, for a good prime p where

$$I(\mathbf{x}_1,\dots,\mathbf{x}_p;\mathbf{y}_p|(y_q)_{q<p})<\varepsilon$$

we note that

$$\mathbb{E}F(\mathbf{x}_1,\dots,\mathbf{x}_p,\mathbf{y}_p)=\mathbb{E}F(\mathbf{x}_1,\dots\mathbf{x}_p,\mathbf{y}'_p)+O(\varepsilon^{1/2}).$$

for any bounded function F where $\mathbf{y}'_p$ is relatively independent of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ over $(\mathbf{y}_q)_{q<p}$. Since $\mathbf{y}_p$ and $(\mathbf{y}_q)_{q<p}$ are already very nearly independent by the Chinese remainder theorem (and in fact if N is a multiple of the product of primes less than p, then $\mathbf{y}_p$ and $(\mathbf{y}_q)_{q<p}$ are genuinely independent) we can conclude that

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p) = \mathbb{E}F(\mathbf{x}_1, \dots \mathbf{x}_p, \mathbf{y}'_p) + O(\varepsilon^{1/2}),$$

where now $\mathbf{y}'_p$ is genuinely independent of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$. For example, if we want to evaluate

$$\frac{1}{N}\sum_{n\leq N}\mu(n)$$

we could interpret this as

$$\mathbb{E}F(\mathbf{x}_0)$$

where $F(x) = x$. Alternately, we can average

$$\frac{1}{N}\sum_{n\leq N}\mu(n) \approx \frac{1}{p}\sum_{i\leq p}\mu(n+i),$$

which is

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p)$$

where now $F(x_1, \dots, x_p) = \frac{1}{p}\sum_{i\leq p} x_i$. Now let $\mathbf{y}'_p$ as before be independent of $(\mathbf{x}_1, \dots, \mathbf{x}_p)$ and uniformly distributed among residue classes mod p. Then this is also

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}'_p)$$

where

$$F(x_1, \dots, x_p, y_p) = \frac{1}{p}\sum_{i\leq p} x_i p \mathbb{1}_{y_p=-i}.$$

As we noted, for p a good prime, this is approximately,

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}'_p) \approx \mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p)$$

and unpacking definitions this is

$$\mathbb{E}F(\mathbf{x}_1, \dots, \mathbf{x}_p, \mathbf{y}_p) = \frac{1}{N}\sum_{n\leq N}\frac{1}{p}\sum_{i\leq p}\mu(n+i)p\mathbb{1}_{n=-1 \bmod p}.$$

Undoing the averaging in i gives

$$\approx \frac{p}{N}\sum_{n\leq N}\mu(n)\mathbb{1}_{p|n}.$$

Thus, the analogue of Corollary 2.4 can be proved using the entropy decrement argument, which can be interpreted in the dynamical setting.

The rest of the proof can also be translated to the dynamical setting. The Furstenberg system corresponding to the Möbius function can be constructed as follows. The underlying space is the set of functions from $\mathbb{Z}$ to $\{-1, 0, 1\}$. We construct a random variable on this space. Consider a random shift of the Möbius function. Formally, let $\mathbf{n}$ be a uniformly chosen random integer between 1 and N and let $\mathbf{X}_N$ denote the function μ (say extended by 0 to the left) shifted by $\mathbf{n}$ i.e. $\mathbf{X}_N(i) = \mu(i+\mathbf{n})$. Since the underlying space of functions from $\mathbb{Z}$ to $\{-1, 0, 1\}$ is compact, there is a subsequence of $(\mathbf{X}_N)_N$ which converges weakly to a random variable $\mathbf{X}$. Since the distribution of each random variable $\mathbf{X}_N$ is "approximately" shift invariant, the distribution of the limit $\mathbf{X}$ is actually shift invariant. Thus, we

obtain a shift invariant measure ν on the space of functions from $\mathbb{Z}$ to $\{-1,0,1\}$ with the property that if f is the "evaluation at zero" map

$$f((a_n)_{n\in\mathbb{Z}}) = a_0$$

then

$$\int f(x)\nu(dx) = \mathbb{E}f(\mathbf{X})$$

is a subsequential limit of terms of the form

$$\frac{1}{N}\sum_{n\leq N}\mu(n).$$

Thus, we can encode questions about the average of μ or more generally shifts like $\mu(n)\mu(n+1)$ in a dynamical way.

In order to take advantage of the fact that μ is multiplicative, we need to impose extra structure on the dynamical systems we associate to μ. This extra structure is implicit in [TT18] and [TT19] and is explicitly described first in [Taoc]. See also [Saw] and [McN]. One key feature of multiplicative functions is that they are statistically multiplicative in the sense that for any $\epsilon_1,\dots,\epsilon_k$ in $\{-1,0,1\}$,

$$\begin{aligned}&\frac{p}{N}\#\{n\leq N\colon \mu(n+pi)=\epsilon_i \text{ for all } i \text{ and } p|n\}\\ =&\frac{p}{N}\#\{n\leq N/p\colon \mu(n+i)=-\epsilon_i \text{ for all } i\}+O\left(\frac{1}{p}\right).\end{aligned}$$

(This holds simply by changing variables and using that μ is multiplicative). For N in some subsequence, we can think of the right hand side as

$$\frac{p}{N}\#\{n\leq N/p\colon \mu(n+pi)=\epsilon_i \text{ for all } i\}\approx\nu\{x\colon f(T^{ip}x)=\epsilon_i\}.$$

We would like a way of encoding this identity in our dynamical system. One solution is to use logarithmic averaging. Now let $\mathbf{n}$ denote a random integer between 1 and N which is not uniformly distributed but which is logarithmically distributed meaning the probability that $\mathbf{n}=m$ is proportional to $\frac{1}{m}$ for $m\leq N$. Let $\mathbf{X}_N(i)=\mu(n+i)$ be a random translate of the Möbius function. Consider the pair $(\mathbf{X}_N,\mathbf{n})$ in the space of pairs of functions from $\mathbb{Z}$ to $\{-1,0,1\}$ and profinite integers. This product space is compact so there is a weak limit $(\mathbf{X},\mathbf{y})$ where $\mathbf{X}$ is a functions from $\mathbb{Z}$ to $\{-1,0,1\}$ and $\mathbf{y}$ is a profinite integer. Let $T(x,y)=(n\mapsto x(n+1),y+1)$. Let ρ be the distribution of $(\mathbf{X},\mathbf{y})$ which is a T-invariant measure on our space. Consider the map I_p on pairs of functions and profinite integers which are 0 mod p which dilates the function by p, multiplies the function by -1 and divides the profinite integer by p i.e.

$$I_p(x,y)=(n\mapsto -x(pn),y/p).$$

For a point (x,y) in our space, let M denote the projection onto the second factor

$$M(x,y)=y.$$

Let f be the "evaluation of the function at 0" function i.e.

$$f(x,y)=x(0).$$

Then the dynamical system has the following properties, where x is always a function from $\mathbb{Z}$ to $\{-1,0,1\}$, p and q are primes and y is a profinite integer:

1. For all p, for all x and y such that $M(x,y) = 0 \bmod p$,
$$I_p(T^p(x,y)) = T(I_p(x,y)).$$
2. For all p and q, for all x and y where $M(x,y)$ is 0 mod pq, we have
$$I_p(I_q(x,y)) = I_q(I_p(x,y)).$$
3. For all p, and for all measurable functions on our space ϕ,
$$\int \phi(x,y)\rho(dxdy) = \int p\mathbb{1}_{M(x,y)=0 \bmod p}\phi(I_p(x,y))\rho(dxdy) + O\left(\frac{1}{p}\right).$$
4. For all p and for all x and y such that $M(x,y) = 0 \bmod p$ we have that
$$f(I_p(x,y)) = -f(x,y).$$

A tuple $(X, \rho, T, f, M, (I_p)_p)$ where (X, ρ, T) is a measure preserving system and satisfying (1) through (4) is a called a dynamical model for μ. Translating our argument over to the dynamical context, there exists some p such that

$$\int f(x,y)\rho(dxdy) \approx \int f(x,y) \cdot p\mathbb{1}_{M(x,y)=0 \bmod p},$$

with an error term which we may make arbitrarily small by increasing p. On the other hand,

$$\begin{aligned}\int f(x,y) \cdot p\mathbb{1}_{M(x,y)=0 \bmod p} &= \int -f(I_p(x,y)) \cdot p\mathbb{1}_{M(x,y)=0 \bmod p} \\ &= -\int f(x,y).\end{aligned}$$

We conclude that

$$\int f = 0,$$

for any dynamical model for μ.

In [Taoc], Tao constructs a dynamical model where

$$\int f \approx \frac{1}{\log N}\sum_{n\leq N}\frac{1}{n}\mu(n)$$

i.e. using logarithmic averaging and the Furstenberg correspondence principle. However using either Corollary 2.4 or a version of the entropy decrement argument, we can argue as follows. Let ρ_N denote the distribution of $(\mathbf{X}_N, \mathbf{n})$ in the space of pairs of functions $\mathbb{Z} \to \{-1, 0, 1\}$ and profinite integers and where $\mathbf{n}$ is a uniformly distributed random integer between 1 and N and $\mathbf{X}_N(i) = \mu(\mathbf{n} + i)$. For any ϵ in S^1 and ϕ, define $\epsilon_*\rho_n$ by

$$\int \phi(x,y)\epsilon_*\rho_N(dxdy) = \int \phi(\epsilon \cdot x, y)\rho_N(dxdy).$$

Choose ϵ_N so that

$$\nu_m = \left(\sum_{n\leq m}\frac{1}{n}\right)^{-1}\sum_{N\leq m}\frac{1}{N}(\epsilon_N)_*\rho_N,$$

satisfies

$$\int f(x,y)\nu_m(x,y) = \left(\sum_{n\le m}\frac{1}{n}\right)^{-1}\sum_{N\le M}\frac{1}{N}\left|\frac{1}{N}\sum_{n\le N}\mu(n)\right|,$$

i.e. ϵ_N is the sign of $\sum_{n\le N}\mu(n)$. Using a version of Corollary 2.4 or the entropy decrement argument, one can prove that for most p (except for a set of logarithmic size at most a constant depending on ε),

$$(I_p)_*(p\mathbb{1}_{M=0 \bmod p}\,\nu_m) \approx \nu_m + O\left(\varepsilon + \frac{\log p}{\log m}\right).$$

By the argument from before (see the proof of Theorem 2.9), this is enough to conclude the prime number theorem.

Acknowledgement. I would like to thank Florian Richter for his patience and conversation. I would also like to thank Terence Tao for including this proof in his class on number theory and for many helpful discussions. I would like to thank Gergley Harcos and Joni Teräväinen for comments on an earlier draft. I would also like to thank Tim Austin, Will Baker, Bjorn Bringmann, Asgar Jamneshan, Gyu Eun Lee, Adam Lott, Clark Lyons, Bar Roytman, Chris Shriver and Will Swartworth for helpful conversations.

References

[Bom71] E. Bombieri, *A note on the large sieve*, Acta Arith. **18** (1971), 401–404, available at https://doi.org/10.4064/aa-18-1-401-404. MR286773

[Dab89] H. Daboussi, *On the prime number theorem for arithmetic progressions*, J. Number Theory **31** (1989), no. 3, 243–254, available at https://doi.org/10.1016/0022-314X(89)90071-1. MR993901

[Erd49] P. Erdös, *On a Tauberian theorem connected with the new proof of the prime number theorem*, J. Indian Math. Soc. (N.S.) **13** (1949), 131–144. MR33309

[EW17] Manfred Einsiedler and Thomas Ward, *Functional analysis, spectral theory, and applications*, Graduate Texts in Mathematics, vol. 276, Springer, Cham, 2017. MR3729416

[GHS19] Andrew Granville, Adam J. Harper, and K. Soundararajan, *A new proof of Halász's theorem, and its consequences*, Compos. Math. **155** (2019), no. 1, 126–163, available at https://doi.org/10.1112/s0010437x18007522. MR3880027

[Gol04] D. Goldfeld, *The elementary proof of the prime number theorem: an historical perspective*, Number theory (New York, 2003), 2004, pp. 179–192. MR2044518

[Gol73] L. J. Goldstein, *Correction to: "A history of the prime number theorem" (Amer. Math. Monthly* **80** *(1973), 599–615)*, Amer. Math. Monthly **80** (1973), 1115, available at https://doi.org/10.2307/2318546. MR330016

[Hil86] Adolf Hildebrand, *The prime number theorem via the large sieve*, Mathematika **33** (1986), no. 1, 23–30, available at https://doi.org/10.1112/S002557930001384X. MR859495

[Kub64] J. Kubilius, *Probabilistic methods in the theory of numbers*, Translations of Mathematical Monographs, Vol. 11, American Mathematical Society, Providence, R.I., 1964. MR0160745

[McN] Redmond McNamara, *Sarnak's conjecture for sequences of almost quadratic word growth*, available at https://arxiv.org/abs/1901.06460.

[Ric] Florian Richter, *A new elementary proof of the prime number theorem*, available at https://arxiv.org/abs/2002.03255.

[Saw] Will Sawin, *Dynamical models for Liouville and obstructions to further progress on sign patterns*, available at https://arxiv.org/abs/1809.03280.

[Sel50] Atle Selberg, *An elementary proof of the prime-number theorem for arithmetic progressions*, Canad. J. Math. **2** (1950), 66–78, available at https://doi.org/10.4153/cjm-1950-007-5. MR33306

[Taoa] Terence Tao, *254a, notes 9 – second moment and entropy methods.*

[Taob] ———, *A banach algebra proof of the prime number theorem*, available at https://terrytao.wordpress.com/2014/10/25/a-banach-algebra-proof-of-the-prime-number-theorem/.

[Taoc] ———, *Furstenberg limits of the Liouville function*, available at https://terrytao.wordpress.com/2017/03/05/furstenberg-limits-of-the-liouville-function/.

[Tao16] ———, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi **4** (2016), e8, 36, DOI 10.1017/fmp.2016.6.

[TT18] Terence Tao and Joni Teräväinen, *Odd order cases of the logarithmically averaged Chowla conjecture*, J. Théor. Nombres Bordeaux **30** (2018), no. 3, 997–1015, available at http://jtnb.cedram.org/item?id=JTNB_2018__30_3_997_0. MR3938639

[TT19] ———, *The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures*, Algebra Number Theory **13** (2019), no. 9, 2103–2150, available at https://doi.org/10.2140/ant.2019.13.2103. MR4039498

[Zag97] D. Zagier, *Newman's short proof of the prime number theorem*, Amer. Math. Monthly **104** (1997), no. 8, 705–708, available at https://doi.org/10.2307/2975232. MR1476753

Redmond McNamara
e-mail: redmondcmcnamara@gmail.com