On elementary proofs of the Prime Number Theorem for arithmetic progressions, without characters.

Andrew Granville * (Institute for Advanced Study, Princeton)

Abstract: We consider what one can prove about the distribution of prime numbers in arithmetic progressions, using only Selberg's formula. In particular, for any given positive integer q, we prove that either the Prime Number Theorem for arithmetic progressions, modulo q, does hold, or that there exists a subgroup H of the reduced residue system, modulo q, which contains the squares, such that $\theta(x,q,a) \sim 2x/\phi(q)$ for each $a \notin H$ and $\theta(x;q,a) = o(x/\phi(q))$, otherwise. From here, we deduce that if the second case holds at all, then it holds only for the multiples of some fixed integer $q_0 > 1$. Actually, even if the Prime Number Theorem for arithmetic progressions, modulo q, does hold, these methods allow us to deduce the behaviour of a possible 'Siegel zero' from Selberg's formula. We also propose a new method for determining explicit upper and lower bounds on $\theta(x,q,a)$, which uses only elementary number theoretic computations.

1. Introduction.

Define $\theta(x) = \sum_{p \le x} \log p$, where p only denotes primes, and for positive integers a and q, let

$$\theta(x;q,a) = \sum_{\substack{p \le x \\ p \equiv a \pmod q}} \log p.$$

The Prime Number Theorem, which states that there are $\sim x/\log x$ primes $p \le x$ (or equivalently, $\theta(x) \sim x$), was first conjectured by Legendre and Gauss near the start of the last century. By the mid-19th century, Čebyčev had shown how to obtain strong upper and lower bounds on $\theta(x)$ and Riemann had outlined an analytic approach to this

* Supported, in part, by NSF grant number DMS–8610730

question. At the end of the century, Hadamard and De La Vallée Poussin finally proved the Prime Number Theorem, their work based on that of Riemann. Mathematicians, such as Bohr, Hardy and Ingham then observed how the Prime Number Theorem is, intrinsically, equivalent to the associated analytic problem, and so asserted that no "elementary" proof was feasible. It thus came as a surprise when, in the late forties, Selberg [Se1] and Erdös [Er1] constructed an "elementary" proof, though Ingham [In] later showed that it is essentially equivalent to the earlier analytic one of De La Vallée Poussin.

The proofs of Selberg and Erdös, which are based on Selberg's formula

$$\theta(x)\log x + \sum_{p\le x} \theta\left(\frac{x}{p}\right)\log p = 2x\log x + O(x), \tag{1.1}$$

are completely combinatorial in nature. To illustrate this, Erdös [E2] showed that, for any sequence S $(= \{1 < p_1 \le p_2 \le \ldots\})$ of real numbers which satisfies (1.1) (where we let $\theta(x)$ be the sum of $\log p$ over those elements p of S that are $\le x$, and similarly take the sum in (1.1) over such p), we have

$$\sum_{p\le x,\ p\in S} \frac{\log p}{p} = \log x + O(1) \quad \text{and} \quad \theta(x) \sim x. \tag{1.2}$$

Selberg [Se2] extended his arguments to arithmetic progressions, using the analogous formula

$$\theta(x;q,a)\log x + \sum_{\substack{p\le x\\ p\nmid q}} \theta\left(\frac{x}{p};q,\frac{a}{p}\right)\log p = \frac{2}{\phi(q)}x\log x + O\left(\frac{x}{\phi(q)}\right), \tag{1.3}$$

proving that $\theta(x;q,a) \sim x/\phi(q)$ whenever $(a,q) = 1$. However, in his proof, he also had to establish that

$$\left|\sum_{p\le x} \left(\frac{D}{p}\right)\ \frac{\log p}{p}\right| \ll\ D \tag{1.4}$$

for any integer $D \neq 0$ (where $\left(\frac{D}{p}\right)$ is the Jacobi symbol), and related it to the question via the law of quadratic reciprocity. Of course (1.4) is equivalent to showing that $L(1,\chi) \neq 0$

for all real quadratic characters χ, a fact that surely does not follow from (1.3). So, herein, we consider the question of what one can actually prove if one is *only* given the formula (1.3):

Theorem 1. *Given only that (1.3) holds, there are two possibilities for the behaviour of* $\theta(x;q,a)$ *as* $x \to \infty$*:*

(I) $\theta(x;q,a) \sim x/\phi(q)$ *whenever* $(a,q)=1$.

(II) There exists a subgroup $H = H_q$ *of the reduced residue system* $(\mathrm{mod}\, q)$*, containing the squares, and of order* $\phi(q)/2$*, such that*

$$\theta(x;q,a) \;=\; \begin{cases} 2x/\phi(q) \;+\; O(x/\phi(q)\log x) & \text{whenever } a \notin H\ , \\ O(x/\phi(q)\log x) & \text{whenever } a \in H \text{ or } (a,q)>1. \end{cases}$$

The proof of a generalization of this Theorem will occupy the bulk of the paper. Here we note a simple consequence:

Corollary 1. *Suppose that (1.3) holds for any positive integer* q*. If Case (II) of Theorem 1 holds for* q_1 *and* q_2*, then it holds for* $\gcd(q_1,q_2)$*. Thus all integers* q *for which Case (II) holds are multiples of some* q_0 *for which Case (II) holds.*

Proof: Define subgroups J_1 and J_2 of the multiplicative group of residues modulo $\ell = \mathrm{lcm}[q_1,q_2]$, so that $a \in J_i$ if and only if $a \pmod{q_i}$ belongs to H_{q_i}. Then, as $\theta(x;\ell,a) \le \theta(x;q_i,a)$, we see that $\theta(x;\ell,a) = o(x)$ whenever $a \in J_1 \cup J_2$: however, as a consequence of Theorem 1, we know that no arithmetic progression modulo ℓ, can have $\theta(x;\ell,a)$ larger than $2x/\phi(\ell)$ asymptotically, and so at least $\phi(\ell)/2$ arithmetic progressions a modulo ℓ must have $\theta(x;\ell,a) \gg_\ell x$. Therefore $|J_1 \cup J_2| \le \phi(\ell)/2 = |J_1| = |J_2|$, which implies that $J_1 = J_2$. Thus the subgroups H_{q_1} modulo q_1 and H_{q_2} modulo q_2, define the same set of integers, and so must actually define a subgroup of the multiplicative group of residues modulo $\gcd(q_1,q_2)$. Case (II) of Theorem 1 then follows for $\gcd(q_1,q_2)$.

This proof may be modified to give the standard result about the scarcity of 'Siegel zeros', as follows: It is well–known that there exists a character modulo q with a 'Siegel

zero' (that is, $L(s,\chi)$ has a real zero $\beta > 1 - c/\log q$, for some sufficiently small constant $c > 0$) if and only if $\theta(x;q,a) \ \le \ \varepsilon \ x/\phi(q)$ for a range of values of $x > q^A$ and any $a \in H_q$, where H_q is the subgroup of the multiplicative residues modulo q on which χ equals 1 (and where $\varepsilon > 0$ and A are suitably fixed) — see [HB] for precise results (for instance, one can take any $\varepsilon > (1-\beta)\log x$).

We shall show that it is impossible to have values $Q \le q_1, q_2 \le Q^2$, $x = Q^B$ and $\varepsilon_1 + \varepsilon_2 < 1$ with B sufficiently large, such that $\theta(x;q_i,a_i) \ \le \ \varepsilon_i \ x/\phi(q_i)$, for each $a_i \in H_{q_i}$, unless $H_{q_1} = H_{q_2}$. For, if not, then let q be the lcm of q_1 and q_2, and consider $\Sigma = \sum \theta(x;q,a)$ over all $a \notin H_{q_1} \cup H_{q_2}$. Clearly

$$\Sigma \ge \ x - \frac{\phi(q_1)}{2}\varepsilon_1 \frac{x}{\phi(q_1)} - \frac{\phi(q_2)}{2}\varepsilon_2 \frac{x}{\phi(q_2)} \ = \ x\left(1 - \ \frac{\epsilon_1+\epsilon_2}{2}\right).$$

On the other hand, Friedlander's proof in [Fr] implies that (1.3) holds uniformly for $x > q^A$ with error term bounded by the main term times $O((\log q/\log x)^{1/2})$. Therefore taking $x = q^B$ in (1.3) and summing over the $\phi(q)/4$ values of $a \notin H_{q_1} \cup H_{q_2}$, we obtain $\Sigma \le \{1/2 + O(1/B^{1/2})\}x$. Comparing the upper and lower bounds for Σ gives the desired contradiction.

Our main result is a little more general than one might expect. The purpose of stating such an 'abstract' generalization of Theorem 1 is to ensure that it is clear what information we use in the proof.

Theorem 2. *Suppose that we are given a finite, abelian group G and an infinite set of real numbers S, together with a map $\alpha : S \mapsto G$ for which $\alpha(ab) = \alpha(a)\alpha(b)$ for all $a, b \in S$, such that, for all $g \in G$ we have*

$$(1.5) \qquad \sum_{\substack{p \le x, \ p \in S \\ \alpha(p) = g}} \log^2 p \ + \sum_{\substack{pq \le x, \ p,q \in S \\ \alpha(pq) = g}} \log p \log q \ = \ \frac{2x\log x}{|G|} \ + \ O(x).$$

Then there are two possibilities for the behaviour of $\theta_g(x) := \sum_{p \le x, \ p \in S, \ \alpha(p) = g} \log p$, as $x \to \infty$:

(I) $\theta_g(x) \sim x/|G|$ *for each* $g \in G$.

(II) *There exists a subgroup* H *of* G*, of rank 2, such that*

$$\theta_g(x) = \begin{cases} 2x/|G| + O(x/\log x) & \text{if } g \notin H \ ; \\ O(x/\log x) & \text{if } g \in H \ . \end{cases}$$

We remark that our proof here bears many similarities to that in [Er1].

It is clear that Theorem 1 follows from Theorem 2, by taking S to be the set of primes that do not divide q, and $\alpha(p) = a$ where a is the residue class that p belongs to, modulo q. We can also prove a result analogous to Corollary 1:

With G, S and α as in Theorem 2, suppose that we are given subgroups K_1 and K_2 of G, and let $K_3 =< K_1, K_2 >$. Define G_i to be the group of cosets of K_i in G, and $\alpha_i(p)$ to be the coset of K_i to which $\alpha(p)$ belongs. By summing (1.5) over the elements g of a given coset of K_i, we see that (1.5) holds for each G_i. If Case (II) of Theorem 2 holds for both G_1 and G_2, then we can use the same argument as in the proof of Corollary 1, to show that Case (II) of Theorem 2 holds for G_3. We deduce

Corollary 2. *If Case (II) of Theorem 2 holds for* G*, then there exists a subgroup* T *of* G*, such that Case (II) of Theorem 2 holds for* G/K *where* K *is a subgroup of* G *if and only if* K *is a subgroup of* T.

We now consider the group character $\chi = \chi_H$ of G, which is really just the characteristic function of H written multiplicatively (and so can certainly be thought of as an elementary object). Explicitly we define it as

$$\chi_H(g) = \begin{cases} 1 & \text{if } g \in H; \\ -1 & \text{if } g \notin H. \end{cases}$$

Thus (II) may be re–phrased as $\theta_g(x) = (1 - \chi_H(g))x/|G| + O(x/\log x)$.

If one wishes to show that (II) of Theorem 1 is impossible, then one must show that a 'significant' number of primes belong to those arithmetic progressions modulo q, that are in H_q. However once one has done so then an immediate (elementary) deduction is that

$L(1, \chi_H) \neq 0$. Thus ruling out (II) for every q, is equivalent to proving that $L(1, \chi) \neq 0$, for all real, quadratic, Dirichlet characters χ. A number of straightforward methods have been proposed to do this, by Gel'fond, Bombieri and others. Selberg's proof in [Se2] is deduced from (1.4); and we discuss this and other proofs in section 3.

It seems likely that Theorem 2 can be used to provide an elementary proof of the Čebatorev density theorem: One starts by using Deuring's method to reduce all cases (abelian and non–abelian) to the case of cyclic extensions, and then one applies a suitably modified version of Theorem 2.

Most recent work on our subject has concentrated on improving the error term in the Prime Number Theorem (see [Di] for an informative review of the work of Wirsing, Bombieri, Diamond and Steinig, and others). A number of different elementary proofs of $\theta(x) \sim x$ have appeared and we discuss some of these in the next section.

In [Sh3], Shapiro showed that the condition $\theta(x; q, a) \sim x/\phi(q)$ for each $(a, q) = 1$ is equivalent (in an elementary way) to

$$M(x; q, a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod q}} \mu(n) \ = \ o(x), \qquad \text{for each } (a, q) = 1.$$

Thus an alternate approach to these questions is through the identity

$$M(x; q, a) \log x \ = \ - \sum_{p \leq x, \ p \nmid q} M(x/p; q, a/p) \log p \ + \ O(x).$$

However it turns out that this does not help us decide between (I) and (II) since we prove in section 9:

Corollary 3. *In case (I) of Theorem 1 we have $M(x; q, a) = o(x)$ for each $(a, q) = 1$. In case (II) of Theorem 1 we have $M(x; q, a) \sim \chi_H(a) \nu_q x/q$, for each $(a, q) = 1$, where*

$$\nu_q = \frac{6}{\pi^2} \prod_{p \in H} \left(1 - \frac{2}{p+1}\right)$$

is a positive constant.

Remark: By definition we see that $\nu_q = 1/\zeta_H(1)$, where $\zeta_H(s) := \zeta(s)L(s,\chi_H)$.

There are already many papers in the literature that provide explicit upper and lower bounds for $\theta(x)$ and $M(x)$ (for instance, by Čebyčev, Ramanujan, Kalmar, Diamond and Erdös, and many others), and some work has been done on $\theta(x;q,a)$. But, in the case of $\theta(x;q,a)$, the bounds make much use of analytic methods (see [Mc]); in particular, requiring information about the zeros of the associated L-functions. We shall exhibit a new method to obtain explicit upper and lower bounds, which uses only counts of primes in the computation. Actually our result works for the more general equation (1.5):

Theorem 3. *Given (1.5), there exists a computable constant $\tau > 0$ such that*

$$\left|\frac{\theta_g(x)}{x/|G|} - 1\right| \le \max_{\substack{x_0 \le y \le x_0^2 \\ h \in G}} \left|\frac{\theta_h(y)}{y/|G|} - 1\right| + \frac{\tau}{\log x_0}, \tag{1.6}$$

for any $x \ge x_0 \ge 2$ and $g \in G$.

(**Remark:** If the explicit dependancies of the 'O' in (1.2) and (1.5) on G are $O(\kappa_G)$ and $O(\kappa_G x)$, respectively, then $\tau = \tau_G \ll 1 + \kappa_G$.)

The following modification, of Theorem 3, will presumably give much better bounds in practice:

Theorem 3'. *Given (1.5), there exist computable constants $\tau, \tau^* > 0$ such that*

$$\left|\frac{\theta_g(x)}{x/|G|} - 1\right| \le \max_{x_0 \le y \le x_0^2} \left\{\left|\frac{\theta(y)}{y} - 1\right| + \frac{1}{|G|}\sum_{h\in G}\left|\frac{\theta_h(y)}{y/|G|} - 1\right|\right\} + \frac{\tau}{\log x_0} + \tau^*\left(\frac{\log\log x}{\log x}\right)^{1/2},$$

for any $x \ge x_0^2 \ge 4$ and $g \in G$.

Equations (1.1) and (1.5) seem, at first sight, somewhat strange criteria to expect of our sequence S. In fact, in analogy with the natural numbers, we can explain why these

are plausible: Given S define $\mathbf{N}$ to be the set of all products of the form $n = \prod_{i \in I} p_i^{a_i}$, where I is any finite subset of the positive integers, and each $a_i \geq 1$. Given α and G define $\alpha(n) = \prod_{i \in I} \alpha(p_i)^{a_i}$, and $\mu(n) = (-1)^{|I|}$ if each $a_i = 1$, and 0 otherwise. Finally define $\mathbf{N}(x)$ to be the number of elements in $\mathbf{N}$ that are $\leq x$, and $\mathbf{N}_g(x)$ to be the number of such n for which $\alpha(n) = g$. If there exist constants $c > 0$ and $\varepsilon > 0$ such that $\mathbf{N}(x) = cx + O(x/\log^{3+\varepsilon} x)$, then by following the proof of the corresponding formulae in [Sh1] and making the obvious modifications, we can prove that (1.1) holds. If, in addition, we know that $\mathbf{N}_g(x) = cx/|G| + O(x/\log^{3+\varepsilon} x)$ for each $g \in G$, then we deduce (1.1), (1.2), and then following the method of proof of (1.3) given in [Se1], we may obtain (1.5).

Of course, the material in the paragraph above leads us to the subject of Beurling's 'generalized prime numbers' [Be]. Indeed Beurling was able to deduce the Prime Number Theorem if $3 + \varepsilon$ above is replaced by $3/2 + \varepsilon$, and gave counterexamples for when it is replaced by $3/2$. A good review of work in this area may be found in [BD].

Our main result here (Theorem 2) may be interpreted as saying that there are two 'stable solutions' for sequences S satisfying (1.5): Either that the images of the sequence (under the map α) are equi–distributed among the elements of G, or else there is a subgroup H of G of rank 2 such that the images of almost all elements of S belong to the complement of H. Theorem 3 gives us some further information: If, for some interval of the form $[x_0, x_0^2]$, we are in the first case, then we remain there thereafter. Thus an interesting possibility arises: Up to some point we are in the second case, and then the situation becomes a little 'unstable', and we 'collapse' in to the first case. This has a direct analogue in understanding the distribution of primes in arithmetic progressions to a moduli q that has a 'Siegel zero':

Suppose that $L(1-\nu, \chi) = 0$, where χ is the primitive real quadratic character modulo q, for some $\nu < c/\log q$ (for some suitably small positive constant $c > 0$). Then, by the explicit formula for primes in arithmetic progressions, one has (see pgs. 54–56 in [Bo2]),

$$\theta(x; q, a) = \frac{x}{\phi(q)} - \chi(a)\frac{x^{1-\nu}}{\phi(q)} + O\left(\frac{x}{\phi(q)\log x}\right)$$

whenever $x > q^A$ (where A is a fixed absolute constant). Thus, if $\nu \log x \to 0$ as $x \to \infty$ then we are in the second case, with almost all primes $\leq x$ belonging to those residue classes $a \pmod q$ with $\chi(a) = -1$; but, as $\nu \log x$ gets larger and moves up to a constant, and then off to ∞, we 'collapse' into the second case, with the primes $\leq x$ equi–distributed among the arithmetic progressions $\pmod q$.

In section 7 we will discuss this phenomenon (somewhat informally), sketching a proof of the following result which shows that the 'collapse' must happen at the same rate as above (if it occurs at all).

Theorem 4. *Assume the hypothesis of Theorem 2. Fix $\varepsilon > 0$. There exists a constant x_G (depending only on the 'level of uniformity' in (1.5)), such that if case (I) of Theorem 2 holds, but $|\theta_a(y) - y/|G|| > \varepsilon y/|G|$ for some $y \geq x_{|G|}$ and $a \in G$, then there exists a subgroup H of G of rank 2, and values $1 > \omega_H > \nu_H > 0$ such that*

$$(1.7) \qquad \frac{x^{1-\omega_H}}{|G|} + o\left(\frac{x}{|G|}\right) \ \leq \ -\chi_H(g)\left\{\theta_g(x) - \frac{x}{|G|}\right\} \ \leq \ \frac{x^{1-\nu_H}}{|G|} + o\left(\frac{x}{|G|}\right),$$

for all $x \geq x_G$ and all $g \in G$.

We would like to replace (1.7) by the statement

$$(1.7)' \qquad \theta_g(x) = \frac{x}{|G|} - \chi_H(g)\frac{x^{1-\nu_H}}{|G|} + o\left(\frac{x}{|G|}\right),$$

for some fixed $\nu_H > 0$. In section 8 we will show that one can deduce (1.7)' from the hypothesis of Theorem 4, provided one can do so for the quotient group G/H; in other words, to prove Theorem 4 with the conclusion (1.7)' for any group G, it suffices to do so just for the group G of two elements. This will lead us to pose, in section 8, a problem of which the solution would solve this and another, related, problem.

We have seen, in Theorem 4, that (1.5) can be used to account for the effect, on estimates for $\theta_g(x)$, of certain zeros of the L–functions corresponding to the group characters of G. However, it is unlikely that (1.5) could be used to account for the effect of other zeros, further to the left in the complex plane, on estimates for $\theta_g(x)$. This is because

their effect, for x larger than a large power of $|G|$, will be $O(x/|G|)$; and so could not be detected by an 'identity' with that same error term (as in (1.5)). Indeed even (suitable) modifications of the much stronger elementary identities of Diamond and Steinig [DS] have too large error terms to overcome this difficulty.

In conclusion, if one were able to establish (1.3) uniformly for $\log x/\log q \to \infty$ then, by plausible strengthenings of the arguments in this paper, and by solving the problem posed in section 8, one might hope to deduce that the following are the *only* possibilities for the behaviour of $\theta(x;q,a)$ in this range:

(Ia) There exists $\varepsilon > 0$ such that

$$\varepsilon x/\phi(q) \ \le \ \theta(x;q,a) \ \le \ (1-\varepsilon)x/\phi(q)$$
$$\text{and} \ \ \theta(x;q,a) \ \sim \ x/\phi(q);$$

(Ib) There exists a subgroup $H = H_q$ of the reduced residue system $(\mathrm{mod}\, q)$ of order $\phi(q)/2$, and a constant $\nu_q > 0$, such that

$$\theta(x;q,a) = \frac{x}{\phi(q)} - \chi_H(a)\frac{x^{1-\nu_q}}{\phi(q)} + O\left(\frac{x}{\phi(q)\log x}\right);$$

(II) There exists a subgroup $H = H_q$ of the reduced residue system $(\mathrm{mod}\, q)$ of order $\phi(q)/2$, such that

$$\theta(x;q,a) \ = \ \left\{1 - \chi_H(a) + O\left(\frac{1}{\log x}\right)\right\} \ \frac{x}{\phi(q)}.$$

Moreover if Case (II) holds for some q, then it only holds for those q that are multiples of some fixed $q_0 > 1$, and (Ib) does not occur at all. If Case (II) never holds then (Ib) cannot hold for two moduli q_1, q_2 in an interval of the form $[Q, Q^2]$, unless it holds for $\gcd(q_1, q_2)$.

Acknowledgements: I'd like to thank Professor Enrico Bombieri with whom I've had numerous conversations on the subject of this paper, which have been a great help in thinking through much of the material presented herein. Thanks are also due to Professor Harold Diamond for making several pertinent remarks.

2. Elementary Proofs of the Prime Number Theorem—A brief survey.

We start this section with a sketch of the original elementary proof:

Define L and U to be, respectively, the lim inf and lim sup of $\theta(x)/x$ as $x \to \infty$. By (1.1),

$$\begin{aligned}\theta(x)\log x &\geq 2x\log x - \sum_{p\leq x}(U+o(1))\,\frac{x}{p}\log p + O(x)\\ &\geq \{2-U+o(1)\}\,x\log x\end{aligned}$$

using Merten's result: $\sum_{p\leq x}\log p/p = \log x + O(1)$. Choosing x large so that $\theta(x) = \{L+o(1)\}x$ we see that $U+L \geq 2$. Now, by (1.1) again,

$$\begin{aligned}\theta(x)\log x &\leq 2x\log x - \sum_{p\leq x}\{L+o(1)\}\frac{x}{p}\log p + O(x)\\ &\leq \{2-L+o(1)\}\,x\log x.\end{aligned}$$

This time choosing x so that $\theta(x) = \{U+o(1)\}x$ we get $L+U \leq 2$ and, together with the above, this gives

$$L+U = 2. \tag{2.1}$$

Now, choose x large so that $\theta(x) = \{U+o(1)\}x$. Then, by (1.1),

$$\sum_{p\leq x}\left(\theta\left(\frac{x}{p}\right) - L\frac{x}{p}\right)\log p = o(x\log x)$$

and so $\theta(x/p) \sim L\ x/p$ for "almost all" primes $p \leq x$ (i.e. those not in a set of primes $p \leq x$ for which $\sum \log p/p = o(\log x)$).

From here the argument runs, essentially, as follows: Choose such a p_o, say with $p_0 \leq x^{\varepsilon}$ and let $x_0 = x/p$. Then, by (1.1), we deduce in a similar fashion that $\theta(x_0/q) \sim U\ x_0/q$ for "almost all" primes $q \leq x_0$. Then, by finding p and q so that x_0/q and x/p are close together, we get a contradiction unless $L = U$, and so $L = U = 1$ by (2.1).

That we can choose such values of p and q was proved by two different, though closely related methods — in [Er1] and [Se1]. Today, Selberg's method, which is less direct, but far less complicated, is the one used by most researchers (see [LV] for a clear exposition), though our work here is based on Erdös's method.

It has long been known that the Prime Number Theorem is equivalent to the assertion that

$$M(x) := \sum_{n \le x} \mu(n) = o(x)$$

This allowed Postnikov, Linnik and others to use the formula

$$M(x)\log x = -\sum_{p \le x} M\left(\frac{x}{p}\right)\log p + O(x) \tag{2.2}$$

in place of (1.1), in their elementary proofs of the Prime Number Theorem. Daboussi [Da1] used the related inequality

$$M(x,y)\log x \le \sum_{p^k \le x, p \le y} M\left(\frac{x}{p^k}, y\right)\log p + \sum_{\substack{n \le x \\ p|n \Rightarrow p \le y}} \log\,(x/n), \tag{2.3}$$

where $M(x,y) = \sum_{\substack{n \le x \\ p|n \Rightarrow p \le y}} \mu(n)$. In both cases the final, and essential step in the argument, is very similar to that of Selberg (for instance, examine the proof of (2)' in [Da1]).

Recently, Hildebrand [Hi] introduced a seemingly new proof, quite unlike anything given previously: He used the large sieve to show that the summatory function of any real valued multiplicative function f of modulus ≤ 1 cannot change value quickly in long "short" intervals. This implies fairly easily that $M(x) = o(x)$.

The best result deduced from (1.1) is claimed by Dusumbetov [Du1], who showed that $\theta(x) = x + O(x/(\log x)^{1-\varepsilon})$, for any fixed $\varepsilon > 0$. He also claimed a result of corresponding strength for $\theta(x; q, a)$ in [Du2].

3. Ruling out case (II) of Theorem 1 — A brief survey.

We wish to show that (II) is impossible in Theorem 1, for every value of q. Thus, for instance, we need only prove that $\sum_{p\le x,\ p\in H_q}(\log p)/p \gg \log x$.

The first, and most elegant method, is due to Selberg [Se2] (we slightly alter his proof to avoid the word 'character'): Let Q be the set of odd, squarefree divisors of q. Using the law of quadratic reciprocity, it is an elementary exercise to show that, for any

$$D \in \begin{cases} (-1)^{(q-1)/2}Q & \text{if 4 does not divide } q, \\ -Q\cup Q & \text{if 4 does divide } q, \text{ but 8 does not divide } q, \\ -2Q\cup -Q\cup Q\cup 2Q & \text{if 8 divides } q, \end{cases}$$

the set $H_D = \{a : (a,q)=1 \text{ and } \left(\frac{D}{a}\right)=1\}$, is a subgroup of $(\mathbf{Z}/q\mathbf{Z})^*$ of rank 2. Moreover each of these subgroups is distinct, and so we have all subgroups of $(\mathbf{Z}/q\mathbf{Z})^*$ of rank 2, by a simple counting argument. Thus the subgroup H in (II) must be precisely the same as H_D for some integer D (which divides $4q$). However the result in (II) then contradicts (1.4).

Selberg obtained (1.4) by evaluating the product

$$N = \prod_{\substack{|u|\le\sqrt{x/2} \\ |v|\le\sqrt{x/2|D|}}} |u^2 - Dv^2|,$$

missing out the term with $u=v=0$, in two different ways (in fact, Selberg did not give (1.4) uniform in D, but the modifications to his argument are straightforward): First he showed, by simple analysis, that $\log N = \frac{2x}{\sqrt{D}}\ (\log x + O(1))$. Second, he evaluated the power to which each prime p divides D: Primes p with $\left(\frac{D}{p}\right) = -1$, only divide those terms $u^2 - Dv^2$ where p divides both u and v, and so do not contribute much to the total. For primes p with $\left(\frac{D}{p}\right) = +1$, Selberg used an elegant lattice point counting method to show that they divide N to the power $4x/p\sqrt{D}$ plus a small error term. He thus deduced that

$$\sum_{p\le x,\ (\frac{D}{p})=1} \frac{\log p}{p} \ = \ \frac{1}{2}\log x \ + \ O(D),$$

and then (1.4) follows from (1.2).

Shapiro gave two different arguments that rule out Case (II), both of which rely on counting primes in algebraic number fields: For a given algebraic number field K, let $\theta_K(x) = \sum_{N\wp \le x} \log(N\wp)$ where $\wp$ is a prime ideal of K and $N\wp$ its norm. In [Sh1], Shapiro verified the identity

$$\theta_K(x)\log x + \sum_{N\wp \le x} \theta_K\left(\frac{x}{N\wp}\right)\log(N\wp) = 2x\log x + O(x);$$

then Erdös's Theorem (that is (1.2)) implies that $\theta_K(x) \sim x$, the *Prime Ideal Theorem*.

Take K to be the qth cyclotomic field. Any prime ideal in K, which does not divide q nor belong to the rational primes, divides a prime $\equiv 1 \pmod q$. Thus, as each such prime splits into $\phi(q)$ prime ideals, $\theta_K(x) \sim x$ implies that $\theta(x;q,1) \sim x/\phi(q)$, which contradicts (II). Shapiro's second approach was to use the easier estimate

$$\sum_{N\wp \le x} \frac{\log N\wp}{N\wp} = \log x + O(1).$$

From this one can easily deduce that $\theta(x;q,1) \gg x/\phi(q)$, contradicting (II).

However both of these approaches rely on proving some quantitative version of the fact that there are $\sim cx$ ideals $\mathfrak{a}$ of K with $N\mathfrak{a} \le x$. As far as I know, this can only be achieved in an 'elementary way' by counting the ideals in each individual ideal class, by counting lattice points in certain ellipses, in a similar spirit to Dirichlet.

Gel'fond, Bombieri, Bateman and others (see [Bo1], for instance) used a method to give lower bounds on $L(1,\chi)$, based on the combinatorial formula

$$\sum_{n\ge 1} \chi(n)\frac{x^n}{(1-x^n)} = \sum_{n\ge 1}\left(\sum_{d|n}\chi(d)\right)x^n,$$

together with the observation $\sum_{d|n}\chi(d) = 1$ if n is a square, and is ≥ 0 otherwise. Thus for $1 > x > 0$ the sum above is $\ge \sum_{n\ge 1} x^{n^2}$, and we obtain a suitable inequality by multiplying through by $(1-x)$ and then choosing x close to 1.

Daboussi's [Da2] elementary proof starts from the inequality

$$|M(x,y;a,q)|\log x \ \le \sum_{\substack{p^k\le x, p\le y\\ p\nmid q}} \left|M\left(\frac{x}{p^k},y;\frac{a}{p^k},q\right)\right|\log p + \sum_{\substack{n\le x, p|n\Rightarrow p\le y\\ n\equiv a(\mathrm{mod}\ q)}} \log\,(x/n)$$

where

$$M(x,y;a,q) = \sum_{\substack{n\le x\\ p|n\Rightarrow p\le y\\ n\equiv a(\mathrm{mod}\ q)}} \mu(n),$$

which generalizes (2.3) in the spirit of (1.3). In his paper, Daboussi avoids the use of characters; however, he quotes a result from [Sh3] that was shown to be equivalent to the non–vanishing of the associated L-series at $s=0$, in [Sh2,II] (and thus at $s=1$, by the functional equation): Specifically, Daboussi's Lemma 7 uses Shapiro's result,

$$\left|\sum_{\substack{n\le x\\ n\equiv a(\mathrm{mod}\ q)}} \frac{\mu(n)}{n}\right| \ = \ O_q(1). \tag{3.1}$$

However, Shapiro (in a remark after the proof of Lemma 4.1 in [Sh3]) writes that this "is equivalent to the statement that $L(0,\chi)\neq 0$ for $\chi\neq\chi_0$, and hence in turn equivalent to the statement that for all $a,\ (a,q)=1$,

$$\sum_{\substack{p\le x\\ p\equiv a(\mathrm{mod}\ q)}} \frac{\log p}{p} \ = \ \frac{1}{\phi(q)}\log x + O(1)."$$

Corradi [Co] makes this much more explicit by showing that if (3.1) holds for *any one* arithmetic progression $b\ (\mathrm{mod}\, q)$ then, for all $(a,q)=1$, we have $M(x;q,a)=o_q(x)$, which was proved to be equivalent to $\theta(x;q,a)\sim x/\phi(q)$ in [Sh3]. Indeed, Corradi generalizes this to any multiplicative function, whose values lie inside or on the unit circle, in place of μ.

4. Some preparatory lemmas.

We shall note here a couple of straightforward consequences of the definitions of S, G and α, and previous results.

Lemma 1. *If (1.5) holds then (1.2) holds.*

To prove this note that by summing (1.5) over all $g \in G$ we obtain (1.1), and then we may apply the results of [Er2], as explained in the introduction, to get (1.2).

Lemma 2. *Equation (1.5) may be re–written in the following different ways:*

$$\theta_g(x)\log x \ + \sum_{p\le x,\ p\in S} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right)\log p \ = \ \frac{2x\log x}{|G|} \ + \ O(x); \tag{4.1}$$

$$\theta_g(x)\log x \ + \ 2\sum_{p\le x^{1/2},\ p\in S} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right)\log p \ = \ \frac{2x\log x}{|G|} \ + \ O(x); \tag{4.2}$$

$$\left(\theta_g(x) - c\frac{x}{|G|}\right)\log x \ + \sum_{p\le x,\ p\in S} \left(\theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) - c'\frac{x}{p|G|}\right)\log p \ = \ O(x); \tag{4.3}$$

$$\left(\theta_g(x) - c\frac{x}{|G|}\right)\log x \ + \ 2\sum_{p\le x^{1/2},\ p\in S} \left(\theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) - c'\frac{x}{p|G|}\right)\log p \ = \ O(x), \tag{4.4}$$

for any c in the range $0 \le c \le 2$, with $c + c' = 2$. Moreover, if H is a subgroup of G of rank 2, and $g \notin H$ then, for $\sigma = \pm 1$,

$$\theta_g(x)\log x \ + \ 2\sum_{\substack{p\le x,\ p\in S\\ \chi_H(\alpha(p))=\sigma}} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right)\log p \ = \ \frac{2x\log x}{|G|} \ + \ O(x). \tag{4.5}$$

These may all be proved from (1.5) using (1.2) to handle any new sums that arise.

Lemma 3. *If (1.5) holds then, for $0 \le y \le x$ and any $g \in G$, we have*

$$\theta_g(x+y) - \theta_g(x) \le \frac{2y}{|G|} + O\left(\frac{x}{\log x}\right). \tag{4.6}$$

This follows from subtracting (4.1) from the same equation with x replaced by $x + y$.

5. Explicit estimates — the proof of Theorem 3.

The proof of Theorem 3: Define, for each $x \geq 1$,

$$\Delta(x) \ = \max_{g \in G} \left| \frac{\theta_g(x)}{x/|G|} - 1 \right| \quad \text{and} \quad \Delta^*(x) \ = \max_{x^\lambda \leq y \leq x} \Delta(y),$$

where $\lambda = 1/\sqrt{2}$. Taking $c = 1$ in (4.4), and choosing $g \in G$ to maximize the absolute value of the first term there, we get

$$\Delta(x) \log x \leq 2 \left\{ \Delta^*(x) \sum_{p \leq x^{1-\lambda}} \frac{\log p}{p} \ + \ \Delta^*(x^\lambda) \sum_{x^{1-\lambda} < p \leq x^{1/2}} \frac{\log p}{p} \ + \ O(1) \right\},$$

(where the sums are over $p \in S$), and so, for $x = y$, we have

$$\Delta(y) \leq (2 - \sqrt{2}) \Delta^*(y) \ + \ (\sqrt{2} - 1) \Delta^*(y^\lambda) \ + \ O\left(\frac{1}{\log x} \right), \tag{5.1}$$

using (1.2) (which is allowed by Lemma 1).

Now, for any y in the range $x \geq y \geq x^\lambda$, we have

$$\Delta^*(y) \leq \max \{ \Delta^*(x), \Delta^*(x^\lambda) \} \quad \text{and} \quad \Delta^*(y^\lambda) \leq \max \{ \Delta^*(x^\lambda), \Delta^*(x^{1/2}) \};$$

and so, by substituting this into (5.1) for each y in this range, we deduce that

$$\Delta^*(x) \leq (2 - \sqrt{2}) \max \{ \Delta^*(x), \Delta^*(x^\lambda) \} \ + \ (\sqrt{2} - 1) \max \{ \Delta^*(x^\lambda), \Delta^*(x^{1/2}) \} \ + \ O\left(\frac{1}{\log x} \right),$$

which implies

$$\Delta^*(x) \leq \max \{ \Delta^*(x^\lambda), \Delta^*(x^{1/2}) \} \ + \ O\left(\frac{1}{\log x} \right).$$

The result follows from an easy induction argument, by taking this equation for $x = x_0^{2\sqrt{2}}, x_0^4, x_0^{4\sqrt{2}}, x_0^8, \ldots$.

Corollary 4. *Let*

$$M = \limsup_{x \to \infty} \max_{g \in G} \left| \frac{\theta_g(x)}{x/|G|} - 1 \right|.$$

If $M \neq 0$ then, for any $x \geq 4$, there exist values y_+, y_-, in the range $x^{1/4} \leq y_+, y_- \leq x$, and g_+, g_-, such that

$$(5.2)_- \qquad \theta_{g_-}(y_-) \ \leq \ (1-M)y_-/|G| \ + \ O(y_-/\log x)$$

$$(5.2)_+ \qquad \text{and} \qquad \theta_{g_+}(y_+) \ \geq \ (1+M)y_+/|G| \ + \ O(y_+/\log x).$$

We thus note that $0 \leq M \leq 1$, and, for whatever value M takes,

$$\liminf_{x\to\infty} \min_{g\in G} \left(\theta_g(x) \Big/ \frac{x}{|G|}\right) \ = \ 1-M \qquad \text{and} \qquad \limsup_{x\to\infty} \max_{g\in G} \left(\theta_g(x) \Big/ \frac{x}{|G|}\right) \ = \ 1+M.$$

Proof: Taking $x_0 = x^{1/2}$ in (1.6), we see that there exists y in the range $x \geq y \geq x^{1/2}$ and $h \in G$ such that

$$\left|\frac{\theta_h(y)}{y/|G|} - 1\right| \ \geq \ M - \frac{2\tau}{\log x};$$

so that either $(5.2)_-$ or $(5.2)_+$ holds. We shall suppose that $(5.2)_+$ holds (the rest of the argument, if $(5.2)_-$ were to hold, is exactly analogous, with the obvious changes of signs and reversal of inequalities), and that $(5.2)_-$ is false; in other words that

$$\theta_g(z) \geq (1-M)z/|G| \ + Cz/\log x,$$

for any fixed $C > 0$, for all $g \in G$, and any z in the range $y \geq z \geq y^{1/2}$. Taking $x = y$ and $c = 1$ in (4.4) we see that the left side of (4.4) is thus

$$\geq \ \frac{y\log y}{|G|}\left(\frac{C|G| - 2\tau + O(M)}{\log x}\right),$$

which contradicts (4.4) if C was chosen sufficiently large.

Corollary 5. *Given the Hypothesis of Theorem 2, suppose we already know that there exists a subgroup H of G, of rank 2, such that $\theta_g(x) = (1-\chi_H(g))x/|G| + o(x)$ for all $g \in G$. Then $\theta_g(x) = (1-\chi_H(g))x/|G| + O(x/\log x)$ for all $g \in G$, and*

$$(5.3) \qquad \sum_{p\in S, \ \alpha(p)\in H} \frac{\log p}{p} \ = \ O(1)$$

Proof: By the hypothesis $M = 1$ (in Corollary 4), and so, by Corollary 4, there exist arbitrarily large values of x for which there is an element $g \in G$ such that $\theta_g(x) = 2x/|G| + O(x/\log x)$: by the hypothesis it is clear that $g \notin H$, if x is sufficiently large. It is also clear from the hypothesis that if y is sufficiently large (say $> x_0$) then for any $h \notin H$ we have $\theta_h(y) \geq y/|G|$. Thus by (4.5) we have

$$\sum_{\substack{p \leq x/x_0, p \in S \\ \alpha(p) \in H}} \frac{\log p}{p} \ \leq \ \frac{1}{x} \sum_{\substack{p \leq x, p \in S \\ \alpha(p) \in H}} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) \log p \ \leq \ O(1).$$

But this is true for arbitrarily large values of x, and so (5.3) follows.

Now, for any $g \notin H$ and $x \geq 1$, (4.1) gives that $\theta_g(x) \leq 2x/|G| + 0(x/\log x) \ll x/|G|$. Thus, by (4.5),

$$\begin{aligned}\left(\frac{2x}{|G|} - \theta_g(x)\right) \ &\leq \ 2 \sum_{\substack{p \leq x, \ p \in S \\ \alpha(p) \in H}} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) \log p \ + \ O(x) \\ &\ll \ x \left(\sum_{p \in S, \ \alpha(p) \in H} \frac{\log p}{p} \ + \ 1 \right) \ \ll \ x,\end{aligned}$$

by (5.3). Thus we have proved that $\theta_g(x) = 2x/|G| + 0(x/\log x)$ for any $g \notin H$.

Finally if $g \in H$ then

$$\sum_{\substack{p \leq x, p \in S \\ \alpha(p) \in H}} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) \log p \ \ll \ x \sum_{p \in S, \ \alpha(p) \in H} \frac{\log p}{p} \ \ll \ x,$$

by (5.3); and, by the above estimate for $\theta_h(x)$ when $h \notin H$,

$$\begin{aligned}\sum_{\substack{p \leq x, p \in S \\ \alpha(p) \notin H}} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) \log p \ &= \ 2 \sum_{\substack{p \leq x^{1/2}, p \in S \\ \alpha(p) \notin H}} \theta_{g\alpha(p)^{-1}}\left(\frac{x}{p}\right) \log p \ + \ O(x) \\ &= \ 2 \sum_{\substack{p \leq x^{1/2}, p \in S \\ \alpha(p) \notin H}} \left(\frac{2x}{|G|} + O\left(\frac{x}{\log x}\right)\right) \ \frac{\log p}{p} \ \ + \ O(x) \\ &= \ \frac{2x \log x}{|G|} + O(x),\end{aligned}$$

where we compute the sum over p using (1.2) and (5.3). Inserting the two estimates above into (4.5) gives $\theta_g(x) = 0(x/\log x)$, which completes the proof of the result.

Sketch of the proof of Theorem 3': If we now define

$$\Delta_g(x) = \frac{\theta_g(x)}{x/|G|} - 1 \quad \text{and} \quad \Delta(x) \ = \ \frac{1}{|G|} \sum_{g\in G} |\Delta_g(x)|,$$

then one can prove (in an almost identical way to Theorem 3), that

$$\Delta(x) \le \max_{x_0\le y\le x_0^2} \ \Delta(y) \ + \ \frac{\tau}{\log x_0},$$

uniformly for all $x_0 \ge 2$. We next show that

$$\Delta_g(x) \ \le \max_{x^{1/2}\le x'\le x} \Delta(x') \ + \ O\left(\left(\frac{\log\log x}{\log x}\right)^{1/2}\right) \tag{5.4}$$

for each $g \in G$, and so complete the proof of Theorem 3':

Let $\delta \asymp (\log\log x/\log x)^{1/2}$, chosen so that there exists an integer J with $(1+\delta)^{2J} = x$; and let $d_j = (1+\delta)^j$ for each j. By (4.4) with $c = c' = 1$ we obtain

$$\begin{aligned} |\Delta_g(x)|\log x \ &\le \ 2\left|\sum_{p\le x^{1/2}} \Delta_{g\alpha(p)^{-1}}(x/p)\ \frac{\log p}{p}\right| \ + \ O(1) \\ &\le 2\sum_{j=0}^{J-1}\left|\sum_{h\in G}\Delta_{gh^{-1}}(x/d_j) \sum_{\substack{d_j<p\le d_{j+1}\\ \alpha(p)=h}} \frac{\log p}{p}\right| \\ &\quad + \ O\left(\sum_{j=0}^{J-1}\left(\sum_{d_j<p\le d_{j+1}}\frac{\log p}{p}\right) \max_{\substack{d_j<p\le d_{j+1}\\ h\in G}} |\Delta_h(x/d_j) - \Delta_h(x/p)|\right). \end{aligned} \tag{5.5}$$

The first term here is

$$\begin{aligned} &\le \ \sum_{j=0}^{J-1}\left(\max_{h\in G}\sum_{\substack{d_j<p\le d_{j+1}\\ \alpha(p)=h}}\frac{\log p}{p}\right)\ \left\{\sum_{h\in G}|\Delta_{gh^{-1}}(x/d_j)| + \left|\sum_{h\in G}\Delta_{gh^{-1}}(x/d_j)\right|\right\} \\ &\le \ \sum_{j=0}^{J-1}\left\{\frac{2}{|G|}\log(1+\delta) + O\left(\frac{1}{j\delta}\right)\right\}\ |G|\ \ \left\{\max_{x^{1/2}\le y\le x}\ \Delta(y) + \left|\frac{\theta(y)}{y} - 1\right|\right\}, \end{aligned}$$

by Lemma 3. The result then follows from this estimate and noting that $|\Delta_h(x/d_j) - \Delta_h(x/p)| \ll \delta$ (by Lemma 3), so that the term in (5.5) is easily bounded.

6. The proof of Theorem 2.

We start by proving

Proposition 1. *Given the hypothesis of Theorem 2, and M as defined in Corollary 4, we have either $M = 0$ or $M = 1$.*

For the rest of this section we shall assume that $M \neq 0$, else we will end up in case (I) of Theorem 2. We will choose arbitrarily small constants $\varepsilon, \delta, \tau, \gamma$, subject to the following constraints:

$$\varepsilon < \frac{M}{5|G|}, \quad \delta \leq \frac{\varepsilon}{2}, \qquad (\text{if } M \neq 1 \text{ then}) \ \tau \leq \frac{\delta}{12} \Big/ \log\left(\frac{2(1+M)}{(1-M)}\right), \quad \gamma \leq \frac{\tau\varepsilon^2}{2|G|^2}.$$

If $M \neq 1$ then we choose them in the order above; if $M = 1$ then we may choose them in the order $\delta, \gamma, \varepsilon = \tau$. We also define $d_j = (1+\delta)^j$ for each $j \geq 0$. With such definitions we may state

Proposition 2. *Suppose that there exists $a \in G$ and $x = d_J$ for some J, such that $\theta_a(x) \geq (1 + M - \gamma)x/|G|$. Then there exists a set $T(= T_{J,a})$ of integers j in the range $\tau J \leq j \leq (1-\tau)J$, of size $\geq (1-3\tau)J$, such that, for each $j \in T$ there exists a set $B_j(= B_{j,J,a})$ of $|G|/2$ elements of G, such that*
If $b \in B_j$ then

$$\theta_{a/b}(d_{J-j-1}) > \left(\frac{1+M}{|G|} - 2\varepsilon\right) d_{J-j-1} \qquad \text{and} \qquad \theta_b(d_{j+1}) - \theta_b(d_j) \leq \frac{2\gamma}{\varepsilon\tau}\ (d_{j+1} - d_j);$$

If $b \notin B_j$ then

$$\theta_{a/b}(d_{J-j-1}) < \frac{(1-M+2\varepsilon)}{|G|} d_{J-j-1} \qquad \text{and} \qquad \theta_b(d_{j+1}) - \theta_b(d_j) \geq \frac{(2-\varepsilon)}{|G|}\ (d_{j+1} - d_j).$$

Proof: Substituting $\theta_a(x) \geq (1+M-\gamma)x/|G|$ into (4.3) with $c = 1+M$ gives

$$\sum_{p\leq x}\left\{\frac{\theta_{a\alpha(p)^{-1}}(x/p)}{x/p|G|} - (1-M)\right\}\frac{\log p}{p} \;\leq\; \gamma\log x + O(1)\,:$$

Now, by Theorem 3, the term inside the brackets here is $\gg -1/\log(x/p)$; therefore there exists a set of integers, T, as described in the hypothesis, for which

$$\sum_{\substack{d_j<p\leq d_{j+1}\\ \alpha(p)=b}}\left\{\frac{\theta_{a/b}(x/p)}{x/p|G|} - (1-M)\right\}\frac{\log p}{p} \;\leq\; \frac{\delta\gamma}{\tau} \tag{6.1}$$

for each $b \in G$. Now define B_j above to be the set of those b for which $\theta_{a/b}(d_{J-j-1}) \geq (1-M+2\varepsilon)d_{J-j-1}/|G|$. For these b, we have

$$\theta_{a/b}(x/p) \;\geq\; (1-M+2\varepsilon)\frac{d_{J-j-1}}{|G|} \;>\; (1-M+\varepsilon)\frac{d_{J-j}}{|G|} \;>\; (1-M+\varepsilon)\frac{x}{p|G|}$$

for all p in the interval $(d_j, d_{j+1}]$. Thus, by (6.1),

$$\sum_{\substack{d_j<p\leq d_{j+1}\\ \alpha(p)=b}}\frac{\log p}{p} \;\leq\; \frac{\delta\gamma}{\varepsilon\tau},$$

and so

$$\theta_b(d_{j+1}) - \theta_b(d_j) \;\leq\; \frac{\delta\gamma}{\varepsilon\tau}d_{j+1} \;\leq\; \frac{2\gamma}{\varepsilon\tau}\,(d_{j+1}-d_j). \tag{6.2}$$

Now by (1.2), and the fact that for any $j \in T$ we have $d_j \geq x^\tau$,

$$\begin{aligned} d_{j+1}-d_j &= \theta(d_{j+1}) - \theta(d_j) + o(d_j)\\ &= \sum_{b\in B_j}\{\theta_b(d_{j+1})-\theta_b(d_j)\} + \sum_{b\notin B_j}\{\theta_b(d_{j+1})-\theta_b(d_j)\} + o(d_j)\\ &\leq |B_j|\frac{\delta\gamma}{\varepsilon\tau}d_{j+1} + \{|G|-|B_j|\}\,\frac{2}{|G|}\,(d_{j+1}-d_j) + o(d_j) \qquad (6.3)\end{aligned}$$

by (6.2) and Lemma 3 respectively. Thus $|B_j| \le |G|/2$. Write d for d_{J-j-1}. Then, by (1.2),

$$\begin{aligned} d + o(d) \;&=\; \theta(d) \;=\; \sum_{b\in B_j} \theta_{a/b}(d) + \sum_{b\notin B_j} \theta_{a/b}(d) \\ &\le\; |B_j|\{1+M+o(1)\}\frac{d}{|G|} \;+\; (|G|-|B_j|)(1-M+2\varepsilon)\frac{d}{|G|} \end{aligned} \tag{6.4}$$

by Corollary 4 and the definition of B_j, respectively. Therefore $|B_j| \ge |G|/2$, and so $|B_j| = |G|/2$. However this implies that the last inequality in each of (6.3) and (6.4) are both, in fact, equalities (asymptotically) and so we get the first and last inequalities in Proposition 2 from (6.4) and (6.3), respectively. This completes the proof of Proposition 2.

Proof of Proposition 1: Suppose that M is not 0 or 1. By Corollary 4, we know that the Hypothesis of Proposition 2 is satisfied for arbitrarily large values of x.

Choose N so that $(1+\delta)^N > 2(1+M)/(1-M)$, so that for $\ge (1-3N\tau)J$ integers $\tau J \le j' \le (1-\tau)J$, each of $j', j'+1, \ldots, j'+N$ are elements of T. We now show

Lemma 4. $B_{j'} = B_{j'+1} = \ldots = B_{j'+N}$.

Proof: If not, there would exist an integer j (with $0 \le j - j' < N$) and $b \in G$, such that $b \in B_j$ but $b \notin B_{j+1}$. Thus, by Lemma 3, and then by Proposition 2,

$$\begin{aligned} \frac{2}{|G|}(d_{J-j-1} - d_{J-j-2}) \;&\ge\; \theta_{a/b}(d_{J-j-1}) - \theta_{a/b}(d_{J-j-2}) \;+\; O\left(\frac{d_{J-j}}{\log x}\right) \\ &\ge\; \left(\frac{1+M}{|G|} - 2\varepsilon\right) d_{J-j-1} - \frac{(1-M+2\varepsilon)}{|G|} d_{J-j-2} \;+\; o(d_{J-j}) \end{aligned}$$

which is impossible.

Continuation of the proof of Proposition 1: If $b \notin B_{j'}$ then, by Lemma 1,

$$\begin{aligned}\frac{(1+M)}{|G|}d_{j'+N} - \frac{(1-M)}{|G|}d_{j'} + o(d_{j'}) &\geq \theta_b(d_{j'+N}) - \theta_b(d_{j'}) \\ &= \sum_{i=0}^{N-1}\{\theta_b(d_{j'+i+1}) - \theta_b(d_{j'+i})\} \\ &\geq \frac{(2-\varepsilon)}{|G|}(d_{j'+N} - d_{j'}).\end{aligned}$$

Dividing this through by $d_{j'}/|G|$, and re–arranging, we obtain
$1 + M - \varepsilon + o(1) \geq (1 - M - \varepsilon)(1+\delta)^N$, which is clearly false, by the definition of N.

We now re–state Proposition 2, given Proposition 1 and that $\gamma \leq \tau\varepsilon^2/2|G|^2$ (by assumption):

Proposition 2. *Suppose that (I) of Theorem 2 does not hold. Then $M = 1$. Now, suppose that we are given $a \in G$ and $x = d_J$ for some J, such that $\theta_a(x) \geq (2-\gamma)x/|G|$. Then there exists a set $T(= T_{J,a})$ of integers j in the range $\tau J \leq j \leq (1-\tau)J$, of size $\geq (1-3\tau)J$, such that, for each $j \in T$ there exists a set $B_j(= B_{j,J,a})$ of $|G|/2$ elements of G, such that*
If $b \in B_j$ then

$$\theta_{a/b}(d_{J-j-1}) > \left(\frac{2}{|G|} - 2\varepsilon\right)d_{J-j-1} \qquad \text{and} \qquad \theta_b(d_{j+1}) - \theta_b(d_j) \leq \frac{\varepsilon}{|G|^2}\ (d_{j+1} - d_j);$$

If $b \notin B_j$ then

$$\theta_{a/b}(d_{J-j-1}) < \frac{2\varepsilon}{|G|}d_{J-j-1} \qquad \text{and} \qquad \theta_b(d_{j+1}) - \theta_b(d_j) \geq \frac{(2-\varepsilon)}{|G|}\ (d_{j+1} - d_j).$$

An easy consequence of this and Corollary 4 is

Corollary 6. *Suppose that (I) of Theorem 2 does not hold, and fix $c > 0$. For any sufficiently large X there exists a value of x $(= d_I$ for some $I)$ in the range $X \le x \le X^5$, and a set A of $|G|/2$ elements of G, such that $\theta_a(x) \ge (2-c)x/|G|$ for each $a \in A$.*

Proof: By Corollary 4, we know that there exists a value of y in the range $X^{5/4} \le y \le X^5$ and $g \in G$, such that $\theta_g(y) = 2y/|G| + O(y/\log y)$. Thus we may satisfy the hypothesis of Proposition 2, with all the constants arbitrarily small, in particular $\varepsilon < c/2|G|$. Selecting any $j \in T$ with $j \le J/10$, our result follows with $I = J - j - 1$ and $A = B_j$, by Proposition 2.

Our next result is

Proposition 3. *Assume that (I) of Theorem 2 does not hold, and $\gamma \ge \delta$ are given, sufficiently small, positive constants (we shall take $\gamma = \delta$ in the proof). Suppose that we are given x $(= d_J$ for some $J)$ and a set A of $|G|/2$ elements of G, such that $\theta_a(x) \ge (2-\gamma)x/|G|$ for each $a \in A$. Then there exists $\eta > 0$ (which is $= \delta^{1/25}$), and a subgroup $H (= H_J)$ of G of rank 2, such that $A = G \setminus H$. Moreover there exists a set $\mathfrak{T}$ $(=\mathfrak{T}_J)$ of integers j in the range $\eta J \le j \le (1-\eta)J$, of size $\ge (1-3\eta)J$, such that, for each $j \in \mathfrak{T}$,*
If $b \in H$ then

$$\theta_b(d_{J-j-1}) < \frac{2\eta}{|G|} d_{J-j-1} \qquad \text{and} \qquad \theta_b(d_{j+1}) - \theta_b(d_j) \le \frac{\eta}{|G|^2}\ (d_{j+1} - d_j);$$

If $b \notin H$ then

$$\theta_b(d_{J-j-1}) > \left(\frac{2}{|G|} - 2\eta\right) d_{J-j-1} \qquad \text{and} \qquad \theta_b(d_{j+1}) - \theta_b(d_j) \ge \frac{(2-\eta)}{|G|}\ (d_{j+1} - d_j).$$

We will prove this after completing the proof of Theorem 2: As a consequence of Proposition 3 we prove

Corollary 7. *Assume the hypothesis of Proposition 3 with $x = d_J$ sufficiently large. There exists a subgroup H $(= H_J)$ of G, of rank 2, such that for any z in the range $x \geq z \geq x^{1/10}$, and any $g \notin H$, we have the lower bound*

$$\theta_g(z) > (2 - 200\eta|G|)\frac{z}{|G|}. \tag{6.5}$$

Proof: Select k so that $d_k \leq z < d_{k+1}$; thus $J/11 \leq k \leq J$. From Lemma 3 we can immediately deduce that if $d_j < p \leq d_{j+1}$ and $k \geq j+2$ then (remembering that δ is fixed)

$$\frac{\theta_g(z/p)}{z/p|G|} \ \leq \ 2 \ + \ O\left(\frac{1}{k-j}\right), \tag{6.6}$$

for any $g \in G$ (and note that $\theta_g(z/p) = 0$ for larger p). Moreover we have

(6.7)
$$\sum_{\substack{d_j < p \leq d_{j+1} \\ \alpha(p) \in H}} \frac{\log p}{p} \ \leq \ \sum_{b \in H} \frac{\theta_g(d_{j+1}) - \theta_g(d_j)}{d_j} \ \leq \ \begin{cases} \delta + O(1/j) & \text{by Lemma 3;} \\ \delta\eta/2|G| & \text{by Proposition 3 for } j \in \mathfrak{T}. \end{cases}$$

Now, from (4.5) with $\sigma = 1$ and then, on the next two lines by (6.6) and (6.7) respectively, we have, for $g \notin H$,

$$\begin{aligned} \left|1 - \frac{\theta_g(z)}{2z/|G|}\right| \log z \ &= \ \sum_{\substack{p \leq z \\ \alpha(p) \in H}} \frac{\log p}{p} \ \frac{\theta_{g\alpha(p)^{-1}}(z/p)}{z/p|G|} \ + \ O(1) \\ &\leq \ \sum_{j=0}^{k-2} \left\{2 + O\left(\frac{1}{k-j}\right)\right\} \sum_{\substack{d_j < p \leq d_{j+1} \\ \alpha(p) \in H}} \frac{\log p}{p} \ + \ O(1) \\ &\leq \ \frac{\delta\eta}{2|G|} \sum_{j=1}^{k-2} \left\{2 + O\left(\frac{1}{k-j}\right)\right\} \ + \ \sum_{1 \leq j \leq k-2}{}' \left\{2\delta + O\left(\frac{1}{j} + \frac{1}{k-j}\right)\right\} \\ &\leq \ \frac{\delta\eta}{|G|}k \ + \ 6\delta\eta|G|J \ + \ O(\log J) \ \leq \ 100\eta|G|\log z, \end{aligned}$$

where $\sum'$ denotes a sum over values of $j \notin \mathfrak{T}$, and the result follows.

Completion of the proof of Theorem 2: Suppose that (I) does not hold and fix positive constants $\gamma \geq \delta$ and $\eta < 1/1000|G|^2$. If X is sufficiently large then take $c = \gamma$ in

Corollary 6, and then, by Corollary 7 (with $J = I$), we know that there exists a subgroup H_X of G such that for all $g \notin H_X$, and for all z in the range $X^{1/5} \geq z \geq X^{1/10}$ (which is guaranteed to be a subrange of $x \geq z \geq x^{1/10}$ as $X \geq x \geq X^{1/5}$), the lower bound (6.5) holds. Thus, for each $g \notin H_X$, we have $\theta_g(z) \geq (2 - 1/5|G|)z/|G|$, and, for each $h \in H_X$, we have

$$\theta_h(z) \ \leq \ \theta(z) - \sum_{g \notin H_X} \theta_g(z) \ < \ z/9|G|,$$

if X is sufficiently large, using (1.2). So we see that each H_X distinguishes those $\theta_g(z)$ that are $< z/5|G|$, from those that are $> 9z/5|G|$: therefore by considering $z = X^{3/20}$ in this way, we see that $H_X = H_{2X}$, and so $H_X = H_{2X} = H_{4X} = \ldots = H$, say.

In the last paragraph we proved that there exists a subgroup H of G, of rank 2, such that if z is sufficiently large then the estimate (6.5) holds, for all $g \notin H$. However, this is true for all fixed $\eta > 0$, and so, as $\theta_g(z) \leq (2 + o(1))z/|G|$ by Lemma 3, we see that $\theta_g(z) \sim 2z/|G|$ for all $g \notin H$. Furthermore, if $h \in H$ then, by (1.2),

$$\theta_h(z) \leq \theta(z) - \sum_{g \notin H} \theta_g(z) = o(z).$$

Thus $\theta_g(x) = (1 - \chi_H(g))x/|G| + o(x)$ for all $g \in G$, and so the result follows from Corollary 5.

The proof of Proposition 3: For now we shall define $\mathfrak{T}_J$ to be the intersection of the sets $T(= T_{J,a})$, as a runs through the set A (where $T_{J,a}$ is as in Proposition 2): for the actual statement of Proposition 3, this definition is amended to the intersection of our set here with the interval $[\eta J, (1 - \eta)J]$. It is then clear, by taking $\delta = \gamma$ and $\eta \geq \tau = \varepsilon$ (and sufficiently large) in Proposition 2, for each $a \in A$, that Proposition 3 follows provided we can prove that $B_j = H$ for each $j \in \mathfrak{T}_J$, and $A = G \setminus H$; thus the remainder of the proof is essentially a complicated exercise in elementary combinatorial group theory. The proof is easiest described in a sequence of steps:

First, taking $\eta > 0$ arbitrarily small (and certainly $< 1/20|G|^2$), define $\delta = \gamma = \eta^{25}$, $\varepsilon = \tau = \eta^8$, $\gamma' = \eta^7$, $\varepsilon' = \tau' = \eta^2$.

Now, for any given $j \in T_{J,a}$, the value of $\theta_b(d_{j+1}) - \theta_b(d_j)$, given by Proposition 2, makes a clear distinction between those b that belong to $B_{j,J,a}$, and those that do not. Thus this set must be the same, no matter what the choice of a and J. In other words,

(i): If $j \in T_{J,a} \cap T_{J',a'}$ then $B_{j,J,a} = B_{j,J',a'}$.
Thus we may denote this set B_j, independent of a and J.

Similarly, for any given $j \in T_{J,a}$, the value of $\theta_{a/b}(d_{J-j-1})$, given by Proposition 2, makes a clear distinction between those group elements that belong to the set $\{a/b : b \in B_j\}$, and those that do not. Thus this set must be the same, no matter what the choice of j, a and J, as long as $J - j$ is fixed. In other words,

(ii): If $j \in T_{J,a}$ and $j' \in T_{J',a'}$, with $J - j = J' - j'$, then $B_j = (a/a')B_{j'}$.

Now, if we take $j = j'$, $J = J'$ and $a, a' \in A$ in (i) and (ii), then we find that $B_j = (a/a')B_j$ for any $a, a' \in A$. Thus $H := \{a/a' : \ a, a' \in A\}$ is a set of $|G|/2$ elements with $A = aH$ and $B_j = b_j H$ for any $a \in A$ and $b_j \in B_j$. But then it is clear that H is a subgroup of G of rank two, as it is finite, abelian, contains 1 and has order $|G|/2$ (all trivially), and is closed (for if $h, h' \in H$ then, for $a' = ah$ and $a'' = a'h'$, we have $hh' = a''/a \in H$). So, we have proved

(iii): Assuming the hypothesis of Proposition 3, we know that there exists a subgroup $H(= H_J)$ of G of rank 2, such that $A = aH$ for any $a \in A$. Moreover, if $j \in \mathfrak{T}_J$ then the conclusions of Proposition 2 hold for $B_j = b_j H$, for any $b_j \in B_j$.

Thus, to complete the proof of Proposition 3, it only remains to prove that $a \notin H$, and $b_j \in H$ for all $j \in \mathfrak{T}_J$.

Now, for any $i \in \mathfrak{T}_J$, it is clear that the hypothesis of Proposition 3 is fulfilled with the same value for δ, but with γ' as above, $A' = (a/b_i)H$ and $J' = J - i - 1$, by (iii). Now suppose that $j' \in \mathfrak{T}_{J'}$ and $j = i + j' + 1 \in \mathfrak{T}_J$. Therefore, by (iii) (applied to J') and then (ii), we have

$$b_{i+j'+1}H = B_{i+j'+1} = (a/(a/b_i))B_{j'} = b_i b_{j'} H.$$

Thus, we have proved,

(iv): If $i,\ i+j+1 \in \mathfrak{T}_J$ and $j \in \mathfrak{T}_{J-i-1}$ then $b_{i+j+1}H = b_i b_j H$.

This is the key to our proof. The idea now is to select values of i and j in (iv), so as to prove that all of these $b_j \in H$. The first step is to show

(v): If $j, j+d, j+2d \in \mathfrak{T}_J$, and we can find k such that $k+j+d+1,\ k+j+2d+1 \in \mathfrak{T}_J,\ k \in \mathfrak{T}_{J-j-d-1} \cap \mathfrak{T}_{J-j-2d-1}$, and $k+d \in \mathfrak{T}_{J-j-1} \cap \mathfrak{T}_{J-j-d-1}$, then $b_{j+2d}H = b_j H$.

From the hypothesis in (v), we can apply (iv) on four occasions, to get

$$b_j b_{k+d} H = b_{j+k+d+1} H = b_{j+d} b_k H, \quad \text{and} \quad b_{j+d} b_{k+d} H = b_{j+k+2d+1} H = b_{j+2d} b_k H.$$

Multiplying these two equations together, and noting that $g^2 \in H$ for any $g \in G$ (by elementary group theory), we obtain (v).

Now, in order to apply (v), we need to guarantee the existence of k that satisfy the hypothesis. If we restrict our search to only those k satisfying

$$\tau' J \ \le\ k+j+d+1 \ \le\ k+j+2d+1 \ \le\ (1-\tau')J,$$

then, by the definition of the sets $\mathfrak{T}$, all but at most $(12+3|G|)\tau' J$ such values of k satisfy the hypothesis of (v). Thus

(v)': Such a k exists if $j+2d \le (1-(15+3|G|)\tau')J$.

Now suppose that $j, j+2r \in \mathfrak{T}_J$, with $j+2r \le (1-(15+3|G|)\tau')J$. First note that we can certainly select s such that each of $j+s, j+2s, j+r+s \in \mathfrak{T}_J$, and $\le (1-(15+3|G|)\tau')J$. But then applying (v) to the triples $j, j+s, j+2s$ and $j+2s, j+r+s, j+2r$ we get $b_j H = b_{j+2s} H = b_{j+2r} H$. In other words (using the definition of $\mathfrak{T}$ in the Proposition, with $\eta \ge (15+3|G|)\tau'$),

(vi): If $i, j \in \mathfrak{T}_J$, have the same parity, then $b_i H = b_j H$.

In other words there exist $g, g' \in G$ such that, for any $j \in \mathfrak{T}_J$, if j is odd then $b_j H = gH$, and if j is even then $b_j H = g'H$.

Now, as $\mathfrak{T}_J$ contains more than three–quarters of the integers $\le J$, we see that there exists $j \le J/2$ such that $j, j+1 \in \mathfrak{T}_J$. We claim that $B_j = B_{j+1}$, for if not then there exists $b \in B_j$, $\notin B_{j+1}$ so that, by Proposition 2,

$$\theta_{a/b}(d_{J-j-1}) - \theta_{a/b}(d_{J-j-2}) \ \ge \ \left(\frac{2}{|G|} - 2\varepsilon\right) d_{J-j-1} - \frac{2\varepsilon}{|G|} d_{J-j-2},$$

which clearly contradicts Lemma 3. Therefore $b_{j+1}H = b_jH$, and so we may take $g' = g$. Moreover, substituting this fact into (iv), we see that $gH = g^2H = H$. Thus we have proved that $B_j = H$ for all $j \in \mathfrak{T}_J$.

Finally by computing the value of the sums

$$\sum_{j=0}^{J-1} \frac{\theta_b(d_{j+1}) - \theta_b(d_j)}{d_j} \ = \ \frac{\theta_b(d_J)}{d_J} \ + \ \sum_{j=0}^{J-1} \frac{\delta\theta_b(d_j)}{d_j} \ ,$$

using Proposition 2 for $j \in \mathfrak{T}_J$ (in the sum on the left side), for $J-j-1 \in \mathfrak{T}_J$ (in the sum on the right side), and Lemma 3 to bound the terms otherwise, it is clear that the set $aH \ne H$, which concludes the proof.

7. Detecting Siegel zero effects — The ideas behind Theorem 4.

Let's suppose that $|\theta_a(x) - x/|G|| \ge (1-\gamma/2)x/|G|$ for some $a \in G$, and sufficiently large $x \ge x_1$. From Theorem 3 we then deduce that for any x_0 in the range $x_2 \le x_0 \le x$, there exists y in the range $x_0 \le y \le x_0^2$ and $b \in G$, such that $|\theta_b(y) - y/|G|| \ge (1-\gamma)y/|G|$.

Next, by the methods of Propositions 2 and 3, and of Corollary 7, we deduce that, corresponding to each such y, we have a subgroup H_y of G, of rank 2, such that for any z in the range $y^{1/6} \le z \le y^{5/6}$, the equation (6.5) holds for any $g \notin H_y$. However (6.5) defines H_y and so, as the intervals, defined for y, overlap, all of the H_y's are the same, that is they equal H, say. Therefore we have shown (writing γ_1 for $\gamma/2$, and γ_2 for $300\eta|G|^2$ in (6.5)),

If $|\theta_a(x) - x/|G|| > (1-\gamma_1)x/|G|$ for some $a \in G$ and $x \geq x_1$ then, for all z in the range $x_2 \leq z \leq x^{5/6}$, we have $|\theta_g(z) - z/|G|| > (1-\gamma_2)z/|G|$ for each $g \in G$. Reversing the logical order here we deduce

Proposition 4. *If there exists a value of $z \geq x_2$ for which $|\theta_g(z) - z/|G|| \leq (1-\gamma_2)z/|G|$ for some $g \in G$, then*

$$\left|\frac{\theta_a(x)}{x/|G|} - 1\right| \leq 1 - \gamma_1$$

for all $x > \max\{x_1, z^{6/5}\}$ and each $a \in G$.

In other words, if $\theta_g(z) \neq o(z/|G|)$ nor $(2+o(1))z/|G|$, for some $g \in G$ and sufficiently large z, then $\theta_b(x)$ will never stray that far from its expected value thereafter. Moreover, we will show that if $\theta_g(z) \neq o(z/|G|)$ nor $(2+o(1))z/|G|$, then $\theta_b(x)$ converges to $x/|G|$ extremely rapidly for $x > z$: The idea will be to establish that if $\gamma_1 < M < M(1+\lambda) \leq 1-\gamma_1$ (for some value of λ depending on γ_1) and

$$\left|\frac{\theta_a(y)}{y/|G|} - 1\right| \leq M(1+\lambda) \tag{7.1}$$

for all $y \geq x_{M(1+\lambda)}$, then

$$\left|\frac{\theta_a(x)}{x/|G|} - 1\right| \leq M \tag{7.2}$$

for all $x \geq x_M = x^3_{M(1+\lambda)}$. As a consequence (iterating these inequalities several times) we deduce that

$$\left|\frac{\theta_a(x)}{x/|G|} - 1\right| \leq \gamma_1 + (1-\gamma_1)(1+\lambda)^{-\log x/\log x_{(1-\gamma_1)}} \leq \gamma_1 + x^{-\nu}, \tag{7.3}$$

for some constant $\nu > 0$. This gives the upper bound in (1.7): the lower bound is proved trivially by taking ω_H sufficiently close to 1.

Proof of (7.3): We need to prove that (7.1) implies (7.2). We shall assume that (7.1) is true and (7.2) is false, so that we have a value of $x \geq x_M$ and $g \in G$ such that $\theta_g(x) > (1+M)x/|G|$ (an equivalent argument works if $\theta_g(x) < (1-M)x/|G|$). We

deduce a contradiction along the lines of Proposition 1, though we have to take a little more care with uniformity: We start by specifying the constants in Proposition 1 (and so Proposition 2 also), in terms of γ_1 and $|G|$:

$$\varepsilon = \gamma_1/5|G|, \quad \delta = \varepsilon/2, \quad \tau = \varepsilon/20\log(8/\gamma_1), \quad \lambda = \gamma_1^2/5000|G|^5\log(8/\gamma_1).$$

Let $c = 1 + M$ in (4.4), so that

$$\sum_{p \le x^{1/2}} \left\{ \frac{\theta_{g\alpha(p)^{-1}}(x/p)}{x/p|G|} - (1-M) \right\} \frac{\log p}{p} \ \le \ O(1).$$

As each term inside the large brackets is $\ge -\lambda M$ by (7.1), we deduce that

$$\sum_{p \le x^{1/2}} \max\left\{0, \ \frac{\theta_{g\alpha(p)^{-1}}(x/p)}{x/p|G|} - (1-M)\right\} \frac{\log p}{p} \ \le \ \frac{M\lambda}{2}\log x \ + \ O(1).$$

Therefore there exists a set of integers, T, containing at least $(1-2\tau)J/2$ values of j, $\tau J/2 \le j \le J/2$, such that

$$\sum_{\substack{d_j < p \le d_{j+1} \\ \alpha(p) = b}} \max\left\{0, \ \frac{\theta_{g/b}(x/p)}{x/p|G|} - (1-M)\right\} \frac{\log p}{p} \ \le \ \frac{M\lambda\log(1+\delta)}{\tau} \ + \ \frac{c}{\log x}$$

for each $j \in T$ and $b \in G$. Given such a j, define B_j to be the set of b for which $\theta_{g/b}(d_{J-j-1}) \ge (1 - M + 2\varepsilon)d_{J-j-1}/|G|$. Imitating the proof of Proposition 2 we obtain

$$\theta_b(d_{j+1}) - \theta_b(d_j) \ \le \ \frac{2M\lambda}{\varepsilon\tau}\,(d_{j+1} - d_j),$$

for $b \in B_j$. For the given values of the constants, the arguments in (6.3) and (6.4) also carry through, so that $|B_j| = |G|/2$; and then we also obtain versions of the remaining two inequalities in Proposition 2:

$$\theta_{g/b}(d_{J-j-1}) \ \ge \ \left(\frac{1+M}{|G|} - \varepsilon - \frac{M\lambda}{2} + o(1)\right) d_{J-j-1} \quad \text{if } b \in B_j,$$

$$\text{and} \quad \theta_b(d_{j+1}) - \theta_b(d_j) \ \ge \ \frac{(2-\varepsilon)}{|G|}\,(d_{j+1} - d_j) \quad \text{if } b \notin B_j.$$

We now follow the rest of the proof of Proposition 1: First we choose N to be the smallest integer $> \frac{3}{\varepsilon}\log\left(\frac{8}{\gamma_1}\right)$, so that $\tau < 1/6N$ and $(1+\delta)^N > 2(1+M+M\lambda-\varepsilon)/(1-M-M\lambda-\varepsilon)$. Therefore we can choose $j' \in T$ with $J/3 \le j' \le J/2$. Lemma 4 follows, and then the rest of the contradiction from using (7.1) to bound both $\theta_b(d_{j'+N})$ and $\theta_b(d_{j'})$.

8. Reducing to the case $|G| = 2$, and an open question.

The reduction: Let's suppose that case (I) of Theorem 2 holds, but that there is some large value of y, as in the hypothesis of Theorem 4. By (7.3), this would be impossible unless we had a subgroup H of G of rank 2, such that, for some wide range of values of x, 'almost all' elements of S have their image in H (under the map α). Thus, if we define

$$\theta_H(x) \;=\; -\sum_{p\le x} \chi_H(p)\log p \;=\; -\sum_{g\in G} \chi_H(g)\theta_g(x),$$

then, in the above 'wide range of values of x', we have $\theta_g(x) \approx (x - \chi_H(g)\theta_H(x))/|G|$. It is thus useful to have the following result:

Theorem 3". *Given (1.5), there exists a computable constant $\tau > 0$ such that*

$$\frac{1}{x}\left|\theta_g(x) - \frac{(x-\chi_H(g)\theta_H(x))}{|G|}\right| \;\le\; \max_{\substack{x_0\le y\le x_0^2\\ h\in G}} \frac{1}{y}\left|\theta_h(y) - \frac{(y-\chi_H(g)\theta_H(y))}{|G|}\right| \;+\; \frac{\tau}{\log x_0},$$

for any $x \ge x_0 \ge 2$ and $g \in G$.

This is proved in a very similar way to Theorem 3, though we need the following identity, which may be obtained by summing $-\chi_H(g)$ times (4.2) over all elements of $g \in G$:

$$\theta_H(x)\log x \;+\; 2\sum_{p\le x^{1/2}} \theta_H\left(\frac{x}{p}\right)\chi_H(p)\log p \;=\; O(x). \tag{8.1}$$

Thus, as a consequence of Theorem 3", and the comments just above that, we see that if we could only prove that $\theta_H(x) = x^{1-\nu_H} + o(x)$, in the range $x \geq x_{|G|}$ then Theorem 4 would follow. But this statement is equivalent to Theorem 4 in the case that we take G' to be the group G/H of two elements. We have thus reduced Theorem 4 to the case where G has two elements.

Let us now look at equations of the form (8.1) a little more closely: For *any* group character χ of G define

$$\theta(x,\chi) \;=\; \begin{cases} \theta(x) - x & \text{when } \chi = \chi_0, \\ \sum_{g\in G}\chi(g)\theta_g(x) = \sum_{p\leq x}\chi(p)\log p & \text{otherwise,} \end{cases}$$

where we write $\chi(p)$ for $\chi(\alpha(p))$, and χ_0 is the 'principal character'. Summing $\chi(g)$ times (4.1) over all $g \in G$ we obtain

$$\theta(x,\chi)\log x \;+\; \int_1^x \theta(x/t,\chi)\mathrm{d}\theta(t,\chi) \;=\; O(x). \tag{8.2}$$

From Lemma 3 we deduce

Lemma 3'. *If χ is a character of G of order m then*

$$|\theta(x+y,\chi) - \theta(x,\chi)| \;\leq\; c_m y \;+\; O\left(\frac{x}{\log x}\right), \tag{8.3}$$

where $c_m = 1$ for $m = 1, 2$ and

$$c_m \;=\; \begin{cases} (2/m)\operatorname{cosec}(\pi/m) & \text{if } m \geq 3 \text{ is even,} \\ (1/m)\cot(\pi/2m) & \text{if } m \geq 3 \text{ is odd,} \end{cases} \qquad < \; 1.$$

To see this note that, by Lemma 3, c_m is the maximum of $|\alpha|$, where $\alpha = a_0 + a_1\zeta^1 + a_2\zeta^2 + \ldots + a_{m-1}\zeta^{m-1}$, over all sets of real numbers $(a_i)_{i=1}^m$ with $0 \leq a_i \leq 2/m$ and $\sum_i a_i = 1$, where $\zeta = e^{2i\pi/m}$. For such an α, it is clear that if the angle subtended between α and ζ^k is less than that subtended between α and ζ^j, and $a_j \neq 0$ then $a_k = 2/m$, else we could take α' with $a_j' = \min\{2/m, a_j + a_k\}$ and $a_k' = \max\{0, a_j + a_k - 2/m\}$, and have $|\alpha'| > |\alpha|$, which is impossible. Thus for even m, such a value of α is given by

$a_0 = a_1 = \ldots = a_{m/2-1} = 2/m,\ a_i = 0$ otherwise. For odd m two possibilities arise: first that $a_0 = a_1 = \ldots = a_{(m-3)/2} = 2/m,\ a_{(m-1)/2} = 1/m,\ a_i = 0$ otherwise, which gives rise to c_m; second that $a_0 = a_{(m-1)/2} = 3/2m,\ a_1 = a_2 = \ldots = a_{(m-3)/2} = 2/m,\ a_i = 0$ otherwise, which gives a value $c_m \cos(\pi/2m)$. Note that in each case $c_m = 2/\pi + O(1/m)$.

Lemma 3' explains why the only significant difficulties in our proofs occur with regard to characters of order 2; it is straightforward to deduce that $\theta(x,\chi) = o(x)$ from (8.2) and (8.3) when $c_m < 1$.

Write $R(x) = \theta(x,\chi)$, when $m = 1$ or 2, so that $R(x)$ is a real–valued function. Equations (8.2) and (8.3) now read as

$$R(x)\log x \ + \ \int_1^x R(x/t)\mathrm{d}R(t) \ = \ O(x), \tag{8.2}'$$

$$|R(x+y) - R(x)| \ \leq \ y \ + \ O\left(\frac{x}{\log x}\right). \tag{8.3}'$$

From the methods of section 6 and 7, we have seen that if $R(x) = -x + O(x/\log x)$ does not occur then $R(x) = o(x)$. This implies both the Prime Number Theorem and our main Theorem here. We would like now to be able to deduce that if $|R(x)| > \varepsilon x$ for $x > x_0$ (where x_0 depends only on the constants implicit in the 'O's in (8.2)' and (8.3)'), then

$$R(x) \ = \ -\frac{x^{1-\nu}}{1-\nu} + o(x) \qquad \text{for all } x \geq x_0, \tag{8.4}$$

for some constant $\nu,\ 1 > \nu > 0$. The methods of the previous section only give the weaker

$$-\frac{x^{1-\nu_1}}{1-\nu_1} + o(x) \ \leq \ R(x) \ \leq \ -\frac{x^{1-\nu_2}}{1-\nu_2} + o(x),$$

for some $1 > \nu_2 > \nu_1 > 0$. A proof of (8.4) seems to be the most interesting open problem that arises from our work here.

9. The proof of Corollary 3.

Let $z = \frac{1}{2}\log x$ and $P = P(z) = \prod_{p\le z,\ p\in H} p$. Note that if n is squarefree then

$$\chi_H(n)\mu(n) = (1-2)^{\#\{p|n : p\in H\}} = \begin{cases} \sum_{d|(n,P)} \mu(d)2^{\omega(d)} & \text{if each prime } p\in H, \text{ that divides } n, \text{ is } \le z,\\ \sum_{d|(n,P)} \mu(d)2^{\omega(d)} + O(1) & \text{otherwise,}\end{cases}$$

where $\omega(d)$ is the number of prime factors of d. Therefore

$$\chi_H(a) \sum_{\substack{n\le x\\ n\equiv a \pmod q}} \mu(n) = \sum_{\substack{n\le x\\ n\equiv a \pmod q\\ n \text{ squarefree}}} \sum_{d|(n,P)} \mu(d)2^{\omega(d)} + O\left(\sum_{\substack{z<p\le x\\ p\in H}} \sum_{\substack{n\le x\\ n\equiv a \pmod q\\ p|n}} 1\right)$$

$$= \sum_{d|P} \mu(d)2^{\omega(d)} \sum_{\substack{m\le x/d,\ (m,d)=1\\ m\equiv a/d \pmod q\\ m \text{ squarefree}}} 1 + O\left(\sum_{\substack{z<p\le x\\ p\in H}} \left(\frac{x}{pq}+1\right)\right), \tag{9.1}$$

writing $n = dm$. We thus need to estimate

$$\sum_{\substack{m\le y,\ \mu(m)\ne 0,\ (m,d)=1\\ m\equiv b \pmod q}} 1 = \sum_{r|d} \mu(r) \sum_{\substack{s\le (y/r)^{1/2}\\ (s,d)=1}} \mu(s) \sum_{\substack{m\le y,\ rs^2|m\\ m\equiv b \pmod q}} 1$$

$$= \sum_{r|d} \mu(r) \sum_{\substack{s\le (y/r)^{1/2}\\ (s,d)=1}} \mu(s) \left\{\frac{y}{qrs^2} + O(1)\right\}. \tag{9.2}$$

Now

$$\sum_{\substack{s\le (y/r)^{1/2}\\ (s,d)=1}} \frac{\mu(s)}{s^2} = \prod_{p\not| d}\left(1-\frac{1}{p^2}\right) + O\left(\frac{1}{(y/r)^{1/2}}\right),$$

and so the main term in (9.2) is

$$\frac{y}{q}\prod_{p|d}\left(1-\frac{1}{p}\right)\prod_{p\not| d}\left(1-\frac{1}{p^2}\right) + O\left(\frac{y^{1/2}}{q}\sum_{r|d}\frac{1}{r^{1/2}}\right),$$

with error term

$$\ll \sum_{r|d} \sum_{s \le (y/r)^{1/2}} 1 \ \ll \ y^{1/2} \sum_{r|d} \frac{1}{r^{1/2}}.$$

Substituting these estimates into (9.2), and then into (9.1) with $y = x/d$ and $b = a/d$, and noting that, by (II), $\sum_{p \in H,\ p>z} 1/p \ll 1/\log z \ll 1/\log\log x$, we get main term

$$\sum_{d|P} \mu(d) 2^{\omega(d)} \frac{x}{qd} \prod_{p|d} \left(1 - \frac{1}{p}\right) \prod_{p \nmid d} \left(1 - \frac{1}{p^2}\right) \ = \ \left\{\nu_q + O\left(\frac{1}{\log\log x}\right)\right\} \frac{x}{q},$$

with error terms

$$\ll \sum_{d|P} 2^{\omega(d)} \left(\frac{x}{d}\right)^{1/2} \sum_{r|d} \frac{1}{r^{1/2}} \ = \ x^{1/2} \prod_{p|P} \left(1 + \frac{2}{p^{1/2}} + \frac{2}{p}\right) \ \le \ x^{1/2}\ e^{cz^{1/2}} \ \ll \ \frac{x}{q\log\log x},$$

and

$$\sum_{\substack{z < p \le x \\ p \in H}} \left(\frac{x}{pq} + 1\right) \ll \frac{x}{q\log\log x} + \frac{x}{\log^2 x} \ll \frac{x}{q\log\log x}$$

if x is sufficiently large. This completes the proof.

References

[BD] BATEMAN, P.T. and DIAMOND, H.G., Asymptotic distribution of Beurling's generalized prime numbers, in "*Studies in Number Theory*", Vol. 6., (Math. Assoc. Amer., Prentice Hall, 1969), 152-210.

[Be] BEURLING, A., Analyse de la loi asymptotique de la distribution des nombres premiers généralisés, I, *Acta Math.*, 68 (1937), 255-291.

[Bo1] BOMBIERI, E., Limitazioni riguardanti somme di caratteri reali e somme di funzioni completamente moltiplicative, *Istituto Lombardo (Rend. Sc.)*, A 94 (1960), 642–649.

[Bo2] BOMBIERI, E., Le grand crible dans la théorie analytique des nombres, *Astérisque*, 18 (1987, 2nd edition), 103pp.

[Co] CORRÁDI, K., A remark on the theory of multiplicative functions, *Acta Sci. Math. (Szeged)*, 28 (1967), 83–92.

[Da1] DABOUSSI, H., Sur le Théorème des Nombres Premiers, *C.R. Acad. Sci. Paris*, 298 (1984), 161–164.

[Da2] DABOUSSI, H., On the Prime Number Theorem for arithmetic progressions, *J. of Number Theory*, 31 (1989), 243–254.

[Dv] DAVENPORT, H., Multiplicative Number Theory, (2nd Ed.) Springer-Verlag, New York, 1980.

[Di] DIAMOND, H.G., Elementary methods in the study of the distribution of prime numbers, *Bull. Amer. Math. Soc.* 7 (1982), 553–589.

[DS] DIAMOND, H.G. and STEINIG J., An elementary proof of the prime number theorem with a remainder term, *Invent. Math.*, 11 (1970), 199–258.

[Du1] DUSUMBETOV, A., An elementary proof of the asymptotic law for the distribution of prime numbers (Russian), *Izv. Akad. Nauk. UzSSR, Ser. Fiz.–Mat. Nauk.*, 2 (1962), 24–31.

[Du2] DUSUMBETOV, A., An elementary proof of the asymptotic law for the distribution of prime numbers in an arithmetic progression (Russian), *Izv. Akad. Nauk. UzSSR, Ser. Fiz.–Mat. Nauk.*, 3 (1963), 5–12.

[Er1] ERDÖS, P., On a new method in elementary number theory which leads to an elementary proof of the Prime Number Theorem, *Proc. Nat. Acad. Sci.*, 35 (1949), 374–384.

[Er2] ERDÖS, P., On a Tauberian theorem connected with the new proof of the Prime Number Theorem, *J. Ind. Math. Soc.*, 13 (1949), 133–147.

[Fr] FRIEDLANDER, J.B., Selberg's formula and Siegel's zero, in "*Recent progress in analytic number theory*", (Durham proceedings, 1979; Academic Press, London–New York, 1981), 15–23.

[HB] HEATH–BROWN, D.R., Siegel zeros and the least prime in an arithmetic progression, *Quart. J. Math. Oxford*, 41 (1990), 405–418.

[Hi] HILDEBRAND, A., The Prime Number Theorem via the large sieve, *Mathematika* 33 (1986), 23–30.

[In] INGHAM, A.E., *Mathematical Reviews*, 10 (1949), 595–596.

[Le] LEVINSON, N., A motivated account of an elementary proof of the Prime Number Theorem, *Amer. Math. Monthly*, 76 (1969), 225–245.

[Mc] McCURLEY, K.S., Explicit estimates for the error term in the Prime Number Theorem for arithmetic progressions, *Math. Comp.* 42 (1984), 265–285.

[Se1] SELBERG, A., An elementary proof of the Prime Number Theorem, *Ann. Math* 50 (1949), 305–313.

[Se2] SELBERG, A., An elementary proof of the Prime Number Theorem for arithmetic progressions, *Can. J. Math* 2 (1950), 66–78.

[Sh1] SHAPIRO, H.N., An elementary proof of the Prime Ideal Theorem, *Comm. Pure Appl. Math*, 2 (1949), 309–323.

[Sh2] SHAPIRO, H.N., On primes in arithmetic progressions, I: *Ann. Math* 51 (1950), 217–230; II: *Ann. Math* 51 (1950), 231–243.

[Sh3] SHAPIRO, H.N., Some assertions equivalent to the Prime Number Theorem for arithmetic progressions, *Comm. Pure Appl. Math* 2 (1949), 293–308.

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA.

Current Address: Department of Mathematics, University of Georgia, Athens, GA 30602, USA.