Prime Number Patterns

Andrew Granville

To Paulo Ribenboim and his friends, the primes.

1. INTRODUCTION. It has long been known that there are infinitely many primes, and that there are infinitely many primes in any arithmetic progression $a, a+d, a+2d, \ldots$ provided $\gcd(a,d) = 1$ and $d \geq 1$. If we ask slightly more involved questions, such as whether there exist infinitely many primes of the form n^2+1, or infinitely many pairs of primes of the form $p, p+2$, then these questions are open, indeed wide open, though computations suggest that the answer to both of these questions is 'yes.' One slight variation on the above theme is to ask for consecutive terms of an arithmetic progression to be prime. Obviously *any* two primes form an arithmetic progression of length two, so the first nontrivial question along these lines is whether there exist many three-term arithmetic progressions of primes, that is, triples of primes of the form a, $a+d$, $a+2d$ where a and d are nonzero integers. In 1933 van der Corput proved that there are indeed infinitely many such triples of (distinct) primes, and there have been several different proofs subsequently. However, the question as to whether one can have four (or more) primes occurring in arithmetic progression infinitely often had seemed far beyond the reach of the current methods of mathematical research until very recently.

In 2005 this situation changed dramatically with the revolutionary paper [**4**] of Ben Green of Cambridge University in England and Terry Tao of UCLA in the United States. Using a panorama of new ideas they showed that for any integer k there are infinitely many k-term arithmetic progressions of primes, that is, there exist infinitely many distinct pairs of nonzero integers a, d such that

$$a, a+d, \ldots, a+(k-1)d$$

are all prime. Their work is based on ideas from many fields: harmonic analysis (and in particular the ideas of Tim Gowers), ergodic theory, additive number theory, discrete geometry, and the combinatorics of set addition. It has helped bring the techniques of the newly defined mathematical discipline, *additive combinatorics*, to a wider audience, leading to spectacular results in graph theory, group theory, and theoretical computer science as well as analytic number theory. This article is not the place to discuss these developments in depth, and indeed we are going to go in a quite different direction.

It is often the case that the conjectures made in mathematics lie just beyond the horizon of what has already been well established. Thus there are many conjectures as to the distribution of primes in various sequences (check out Paulo Ribenboim's charming book *The little book of* BIGGER *primes* [**9**], or Chris Caldwell's *The Prime Pages* [**3**], a website with a cornucopia of questions, data and discussion). However, the horizon was extended so far by the wonderful breakthrough of Green and Tao that we now have little idea of what lies just beyond it. It had been my purpose to ask some new "beyond the horizon" questions, but something surprising happened: I found that the first few questions I asked myself, which had seemed to be far beyond the Green-Tao theorem, turned out to be simple deductions from the Green-Tao theorem! Let me show you, though first let us examine the Green-Tao theorem in a little more detail.

2. PRIME NUMBER PATTERNS: RESULTS, EXAMPLES, AND PREDICTIONS. Following up on the wonderful breakthrough of Green and Tao [**4**] there are many cute types of patterns of primes that we can now prove to exist. I am interested in trying to find examples of each of these patterns, finding the smallest examples of such patterns,[1] and even predicting how large the smallest example is, in some generality. This leaves lots of challenges for the computationally minded.

2.1. Arithmetic progressions of primes. The smallest arithmetic progression of ten primes is given by 199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089, which we can write as $199 + 210n$, $0 \le n \le 9$. The smallest examples of k-term arithmetic progression of primes, with k between 3 and 21, are given by (though see [**1**] for more details):

Length k	Arithmetic Progression ($0 \le n \le k-1$)	Last Term
3	$3 + 2n$	7
4	$5 + 6n$	23
5	$5 + 6n$	29
6	$7 + 30n$	157
7	$7 + 150n$	907
8	$199 + 210n$	1669
9	$199 + 210n$	1879
10	$199 + 210n$	2089
11	$110437 + 13860n$	249037
12	$110437 + 13860n$	262897
13	$4943 + 60060n$	725663
14	$31385539 + 420420n$	36850999
15	$115453391 + 4144140n$	173471351
16	$53297929 + 9699690n$	198793279
17	$3430751869 + 87297210n$	4827507229
18	$4808316343 + 717777060n$	17010526363
19	$8297644387 + 4180566390n$	83547839407
20	$214861583621 + 18846497670n$	572945039351
21	$5749146449311 + 26004868890n$	6269243827111

The k-term arithmetic progression of primes with smallest last term.

Can we predict, without data, the size of the last term of the smallest k-term arithmetic progression of primes? In order to get a good understanding of this we first seek a formula for how many arithmetic progressions $a, a + d, \ldots, a + (k-1)d$ of primes there are with each prime $\le x$. In Section 4 we will analyze this question in some detail, not just for primes in arithmetic progression but for all the patterns that arise in this article. From a careful analysis of the formulas that arise in Section 4 we expect that the smallest k-term arithmetic progression of primes has largest prime around

$$(e^{1-\gamma}k/2)^{k/2}. \tag{2.1}$$

(Here e is the base of the natural logarithm, and $\gamma = .5772156649\ldots$ is the Euler-Mascheroni constant defined by $\gamma = \lim_{N\to\infty}\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}\right) - \log N$. Most important in this article is that $e^{-\gamma} = .5614594836\ldots$.) In fact, if the largest prime

[1] By "smallest" we mean the example in which the largest prime in the set is smallest (and, if there is a tie, the set in which the second largest prime is smallest, etc.).

of the smallest example of a k-term arithmetic progression of primes is divided by $(e^{1-\gamma}k/2)^{k/2}$, then the quotient lies in $(2/5, 2)$ for each n, $15 \le n \le 21$, a remarkably good fit for such a fast growing function.

Green and Tao were able to show that there exists a k-term arithmetic progression of distinct primes all at most

$$2^{2^{2^{2^{2^{2^{100k}}}}}},$$

a spectacular achievement. Based on (2.1) and the numerical data above we conjecture that this bound should be improvable to $k! + 1$, for each $k \ge 3$.

2.2. Generalized arithmetic progressions of primes. *Generalized arithmetic progressions* (GAPs) are sets of integers of the form

$$a + n_1 b_1 + n_2 b_2 + \cdots + n_d b_d, \tag{2.2}$$
$$\text{with } 0 \le n_1 \le N_1 - 1,\ 0 \le n_2 \le N_2 - 1, \ldots, 0 \le n_d \le N_d - 1,$$

for given integers $a, b_1, b_2, \ldots, b_d$, and integers $N_1, N_2, \ldots, N_d \ge 2$. These appear prominently in the theory of set addition mentioned above. (The GAP in (2.2) has *dimension* d, with *volume* $N_1 \ldots N_d$.) The integers in a GAP are not necessarily distinct, though must be so if there is no linear dependence amongst the b_j with small coefficients.

We are interested in finding generalized arithmetic progressions of distinct primes of any given dimension and volume. Although this appears to be a big generalization of primes in arithmetic progressions (the dimension 1 case), it turns out that such GAPS are easily shown to exist as a consequence of the dimension 1 case:

Let $N = \max_{1 \le j \le d} N_j$ and $k = N^d$. Suppose that we have a k-term arithmetic progression of primes, $a + jq$, $0 \le j \le k - 1$. Let $b_i = N^{i-1}q$ for each i, so that

$$a + n_1 b_1 + n_2 b_2 + \cdots + n_d b_d = a + jq$$

where we write $j = n_1 + n_2 N + n_3 N^2 + \cdots + n_d N^{d-1}$ in base N. Therefore the GAP is a subset of our k-term arithmetic progression (and no two elements of the GAP are equal since each j has a unique representation in base N), and so the GAP is made up entirely of distinct primes, as desired.

The smallest 2-by-k GAPs (i.e., GAPs of the form $a + bi + cj$, $0 \le i \le 1$, $0 \le j \le k - 1$) are:

k	GAP	Last Term
2	$3 + 8i + 2j$	13
3	$7 + 24i + 6j$	43
4	$5 + 36i + 6j$	59
5	$11 + 96i + 30j$	227
6	$11 + 42i + 60j$	353
7	$47 + 132i + 210j$	1439
8	$199 + 3300i + 210j$	4969
9	$199 + 3300i + 210j$	5179

The 2-by-k GAPs of distinct primes with smallest last term

A few other examples of smallest GAPs are:

5	17	29
47	59	71
89	101	113

29	41	53
59	71	83
89	101	113

The 3-by-3 GAPs $5 + 12i + 42j$ *and* $29 + 12i + 30j$.

11	47	83
101	137	173
191	227	263
281	317	353

503	1721	2939	4157
863	2081	3299	4517
1223	2441	3659	4877
1583	2801	4019	5237

The 4-by-3 GAP $11 + 90i + 36j$*, and the 4-by-4 GAP* $503 + 360i + 1218j$.

We have been unable to find a 3-by-3-by-3 GAP of distinct primes.

We expect that the smallest N_1-by-N_2-by-$\cdots$-by-N_d GAP of distinct primes has largest prime around

$$\left(e^{1/\kappa-\gamma}\,\frac{k}{(d+1)}\right)^{k/(d+1)}, \tag{2.3}$$

where $\kappa := N_1N_2\ldots N_{d-1}$, $k := \kappa N_d$, and $N_1, N_2, \ldots, N_{d-1} \le N_d$.

2.3. Balog cubes. In [**2**] Balog proved that there are infinitely many 3-by-3 squares of distinct primes where each row and each column forms an arithmetic progression. Similarly he proved that there are infinitely many 3-by-3-by-3 cubes of distinct primes where each row and each column and each vertical line forms an arithmetic progression. And indeed the same is true in d dimensions.

Balog's gorgeous result is now improved by the result of Green and Tao since any GAP of distinct primes with dimension d and $N_1 = N_2 = \cdots = N_d = N$ gives rise to an N-by-N-by-$\cdots$-by-N Balog cube of primes. Note that a Balog cube does not have to be a GAP; indeed the smallest examples are not:

11	17	23
59	53	47
107	89	71

83	131	179	227
251	257	263	269
419	383	347	311
587	509	431	353

The smallest 3-by-3 and 4-by-4 Balog cubes of primes.

It is not difficult to find many examples of 3-by-3 and 4-by-4 Balog cubes, though I have not yet found a 5-by-5 Balog cube.

Going to the next dimension is also difficult, though we eventually have been able to compute many 3-by-3-by-3 Balog cubes, the smallest of which is:

47	383	719
179	431	683
311	479	647

149	401	653
173	347	521
197	293	389

251	419	587
167	263	359
83	107	131

A 3-by-3-by-3 Balog cube of primes.

(Remember, each row and each column of each 3-by-3 square, which is a layer of the 3-by-3-by-3 Balog cube, is an arithmetic progression of primes, and also the (i, j)th elements of each of the three 3-by-3 squares form an arithmetic progression of primes for each fixed $1 \le i, j \le 3$: for example, for $i = 2$, $j = 1$ we have 179, 173, and 167.) Notice that this is not a GAP, and indeed we have been unable to find a 3-by-3-by-3 GAP of distinct primes.

2.4. Sets of primes, averaging in pairs. Balog [**2**] established yet another remarkable result: there exist arbitrarily large sets A of distinct primes such that for any $a, b \in A$ the average $\frac{a+b}{2}$ is also prime (and all of these averages are distinct). (In Section 3 we indicate why we believe that there exist infinitely large such sets A.) Balog's result also follows from the result of Green and Tao [**4**]:

Suppose that we want A to have n elements. If we did not mind whether the averages were all distinct then we could take any $k(= 2n)$-term arithmetic progression of primes $a + jd$, $0 \le j \le k - 1$, and let $A = \{a + 2jd : 0 \le j \le n - 1\}$. In this case $\frac{1}{2}((a + 2id) + (a + 2jd)) = a + (i + j)d$ is prime, since whenever $0 \le i, j \le n - 1$ we have $0 \le i + j < k - 1$. However, we do want all the averages to be distinct, and so we modify this construction using a Sidon sequence:

A sequence of integers $b_1 < b_2 < \cdots < b_n$ is called a *Sidon sequence* if all of the sums $b_i + b_j$, $i < j$, are distinct. The easiest example is $b_i = 2^i$, in which case the $(b_i + b_j)/2$ give rise to distinct binary expansions, and so distinct integers. (There has been a lot of research on the minimal length $(b_n - b_1)$ of a Sidon sequence with n elements, which, as one might guess, is something like a constant times $\sqrt{n}$.) If we suppose that $b_1 = 0$ (which can be achieved, without loss of generality, by adding a constant to each element of the sequence) then take a k $(= 2b_n)$-term arithmetic progression of primes $a + jd$, $0 \le j \le k - 1$, and the averages of the elements of the set $A = \{a + 2b_i d : 1 \le i \le n\}$ are distinct primes, since the averages take the form $a + (b_i + b_j)d$.

n	Set of primes
2	3, 7
3	3, 7, 19
4	3, 11, 23, 71
5	3, 11, 23, 71, 191
6	3, 11, 23, 71, 191, 443
7	5, 17, 41, 101, 257, 521, 881
8	257, 269, 509, 857, 1697, 2309, 2477, 2609
9	257, 269, 509, 857, 1697, 2309, 2477, 2609, 5417
10	11, 83, 251, 263, 1511, 2351, 2963, 7583, 8663, 10691
11	757, 1009, 1117, 2437, 2749, 4597, 6529, 10357, 11149, 15349, 21757
12	71, 1163, 1283, 2663, 4523, 5651, 9311, 13883, 13931, 14423, 25943, 27611

Sets of n primes whose pairwise averages are all distinct primes, with smallest largest element.

Balog gave a beautiful geometric interpretation of such sets A. Think of the elements of A as labels for the vertices of an $(n - 1)$-dimensional tetrahedron, with the

average of any two elements being placed on the edge in between the associated vertices. Then along any edge of the $(n-1)$-dimensional tetrahedron one can read off a 3-term arithmetic progression of primes.

In Section 4 we will indicate why we expect that the smallest such set A with n elements should have largest prime around

$$(e^{-\gamma}n/2)^{n/2}. \tag{2.4}$$

2.5. Sets of primes, averaging all subsets. Now we want arbitrarily large sets of integers A such that the average of the elements of *any* nontrivial subset S of A is also a prime, and that all of these primes are, in fact, distinct.

For a set A of integers and nontrivial subset S of A, let μ_S be the average of the values in S. If we did not mind whether the μ_S were all distinct then we could take any $k\ (= n(n!))$-term arithmetic progression of primes $a + jd$, $0 \le j \le k-1$, and let $A = \{a + j(n!)d : 0 \le j \le n-1\}$. Then for any nonempty subset J of $\{1, 2, \dots, n\}$, and corresponding $S = S_J = \{a + j(n!)d : j \in J\}$, we have

$$\mu_S = \frac{1}{|J|}\sum_{j\in J}(a + j(n!)d) = a + d\left(\sum_{j\in J} j\right)\frac{n!}{|J|},$$

which is an element of our arithmetic progression and thus prime since $n!/|J|$ is an integer and $0 \le n!\sum_{j\in J} j\big/|J| < n(n!) = k$.

However we want all of the averages to be distinct, and so we modify this construction using *any* set $B = \{b_1 < b_2 < \cdots < b_n\}$ of integers for which the averages μ_S, $S \subset B$, $S \ne \emptyset$, are all distinct. We take any $k\ (= (b_n - b_1)n!)$-term arithmetic progression of primes $a + jd$, $0 \le j \le k-1$, and then let $A = \{a + (b_j - b_1)(n!)d : 1 \le j \le n\}$; the result then follows by working through an argument analogous to that in the previous paragraph.

I am not sure of the simplest construction of such a set B, but one can certainly take $B = \{(j+1)! : 1 \le j \le n\}$. In this case, we can determine each subset S from its average μ_S, and therefore the averages must be distinct: To do so we select J to be the minimal integer for which $(J+1)! \ge \mu_S$ and then r to be the minimal integer for which $r\mu_S \ge (J+1)!$. Then $|S| = r$ and $r\mu_S$ is a sum of distinct factorials, which are easily identified.[2]

The minimal examples are:

n	Minimal set of primes
2	3, 7
3	7, 19, 67
4	5, 17, 89, 1277
5	209173, 322573, 536773, 1217893, 2484733

Sets of n primes all of whose subsets have averages that are distinct primes.

We predict that the smallest example of such a set of n primes will have largest prime of size about

$$2^{2^n}/(e^{\gamma}n)^{2^n/n}. \tag{2.5}$$

[2]To prove all this, first note that if $(J+1)!$ is the largest element of S then $\mu_S \ge \min_{i\le J}(2! + 3! + \cdots + i! + (J+1)!)/i \ge (J+1)!/J > J!$ and $\mu_S \le \max_{i\le J}((J+2-i)! + \ldots! + (J+1)!)/i \le (J+1)!$. Now $r \le |S| \le J$, and so if $r = J$ then $|S| = J = r$. Otherwise $r \le J-1$ and so $(r+1)\mu_S = (1+1/r)r\mu_S \ge (1+1/(J-1))(J+1)! > (J+1)! + J! + \cdots + 2!$ so $|S| < r+1$, and therefore $|S| = r$.

I had been unable to find examples with more than 4 elements! Tony Noe sent me the 5-element example above on August 8, 2006. If (2.5) is correct then 6-element examples will be hard to find whereas 7-element examples will remain out of reach.

2.6. Initial polynomial values that are primes. The arithmetic progression $a + jd$, $0 \le j \le k-1$, can equally be thought of as the first k values of the polynomial $dX + a$. Thus the Green-Tao theorem tells us that there are infinitely many linear polynomials for which the first k values are prime. This then suggests the question, how about quadratic polynomials and beyond?

The infamous quadratic polynomial $X^2 + X + 41$ is prime for $X = 0, 1, \dots, 39$, and it is known that 41 is the largest integer m for which $X^2 + X + m$ is prime for $X = 0, 1, \dots, m-2$. This is "equivalent" to the classification of imaginary quadratic fields of class number one, a famously difficult result. More simply one can ask whether, for each k, there exists m such that $X^2 + X + m$ is prime for $X = 0, 1, \dots, k$. We believe so, as we will explain in Section 3.6b.

We have so far taken our prime producing quadratic polynomials in a particular form, most importantly monic. If we drop this requirement then the examples known with the most initial prime values are (the absolute values of) $36X^2 - 810X + 2753$ and $36X^2 - 2358X + 36809$, both of which give distinct primes for $X = 0, 1, \dots, 44$. Notice that these both have discriminant $259668 = 6^2 \times 7213$ and that p is not congruent to a square (mod 7213) for any prime p in the range $5 \le p \le 53$, so no prime ≤ 53 ever divides values of these polynomials.

For cubic polynomials there are several nice examples: The absolute value of the polynomial $(X-14)^3 + (X-14)^2 + 17$ is prime for all integers X, $0 \le X \le 24$. The best polynomial known of degree 3 is $|3n^3 - 183n^2 + 3318n - 18757|$, which is prime for $0 \le n \le 46$; note that no prime < 37 ever divides the value of this polynomial.

Following a 2006 programming competition [**10**], we now know that $|X^4 - 97X^3 + 3294X^2 - 45458X + 213589|$ is prime for all integers X, $0 \le X \le 49$; and that $\frac{1}{4}|X^5 - 133X^4 + 6729X^3 - 158379X^2 + 1720294X - 6823316|$ is prime for all integers X, $0 \le X \le 56$.

More generally we can ask whether there are infinitely many polynomials of degree d which give distinct primes for $X = 0, 1, \dots, k$. Balog [**2**] showed that this is so for $k = 2d$ (generalizing the "three primes in arithmetic progressions" theorem, which is the case $d = 1, k = 2$). The general result follows easily from the Green-Tao theorem, for if we take a $(k^d + 1)$-term arithmetic progression of primes $a + jb$, $0 \le j \le k^d$, then the polynomial $bX^d + a$ is prime for $X = 0, 1, 2, \dots, k$.

We have been unable to settle the more interesting question as to whether there are infinitely many monic polynomials of degree d for which the first k values give distinct primes.

We predict that if $d \ge 2$ and k is large compared to d (in fact $k \ge 4(d \log d)^2$) then there exists a monic polynomial $f(X)$ of degree d, with all of its coefficients positive, for which each of $f(0) < f(1) < \cdots < f(k-1)$ are prime, where $f(k-1)$ is about

$$\left(e^{-\gamma}\,\frac{k}{d}\right)^{k/d}. \tag{2.6}$$

If we allow the leading coefficient of f to be any nonzero integer then we make a similar prediction, adjusting (2.6) by replacing d with $d + 1$.

2.7. Monochromatic arithmetic progressions of primes. Green and Tao actually proved a substantially stronger theorem than that used above: Fix any $\delta > 0$ and any

integer $k \geq 3$. If x is sufficiently large (that is, x is larger than a constant that depends only on δ and k) and if P is a subset of the primes up to x containing at least $\delta\pi(x)$ elements[3] then P contains a k-term arithmetic progression of primes. One easy consequence of this is that if you color the primes with r colors then there will be monochromatic[4] arithmetic progressions of primes of arbitrary length: For if we want an arithmetic progression of length k then take $\delta = 1/r$ in the result above with x sufficiently large. Let $P_1, \ldots, P_r$ be the partition of the primes up to x into their assigned colors. At least one of the P_j has $\geq \delta\pi(x)$ elements, so contains a k-term arithmetic progression of primes of color j by the (strong) Green-Tao theorem.

2.8. Arithmetic progressions of primes, beginning with a given prime. We expect that for any given prime p there are infinitely many k-term arithmetic progressions of primes which begin with p, for any $k \leq p$ (that is, an arithmetic progression of primes $p + jd$, $j = 0, 1, \ldots, k-1$). It is evident that we cannot have $k \geq p+1$ because the $(p+1)$th term of the arithmetic progression, $p + pd = p(1+d)$, will be composite.

Terry Tao pointed out to me that the strong version of the Green-Tao theorem (given in Section 2.7) implies that, for any fixed k, "almost all" primes are indeed the first term of a k-term arithmetic progression of primes. More precisely, suppose that there are $(1 - \epsilon_k(x))\pi(x)$ primes up to x which are the first term of a k-term arithmetic progression of primes. We are claiming that for each fixed k, we have $\epsilon_k(x) \to 0$ as $x \to \infty$.

Fix $\delta > 0$ arbitrarily small and suppose that x is sufficiently large (as in Section 2.7). Let Q be the set of primes $\leq x$ which are the first term of some k-term arithmetic progression of primes all $\leq x$, and let P denote the rest of the primes. Note that P cannot contain a k-term arithmetic progression of primes, else the first term of the arithmetic progression would be in Q not P. Therefore $|P| \leq \delta\pi(x)$ (by the strong version of the Green-Tao theorem given in Section 2.7) and so $|Q| \geq (1-\delta)\pi(x)$. Our claim follows.

We predict that the smallest k-term arithmetic progression of primes that begins with p will have largest term around

$$\left(e^{1-\gamma} \max\left\{k, \frac{\log p}{\log k}\right\}\right)^{k-1}; \tag{2.7}$$

in particular for $k = p$ this yields $(e^{1-\gamma} p)^p$. Computations have yielded the following data:

p	Arithmetic Progression	Last Term
2	$2 + n$	3
3	$3 + 2n$	7
5	$5 + (2 \cdot 3)n$	29
7	$7 + 5 \cdot (2 \cdot 3 \cdot 5)n$	907
11	$11 + 7315048 \cdot (2 \cdot 3 \cdot 5 \cdot 7)n$	15361600811
13	$13 + 4293861989 \cdot (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11)n$	119025854335093
17	$17 + 1138781900732 5752 \cdot (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13)n$	5471619276639877320977

The smallest p-term arithmetic progression of primes beginning with p.

[3]Here and throughout, $\pi(x)$ denotes the number of primes $\leq x$.

[4]That is, all of the same color.

(Note that the common difference must be divisible by every prime $q < p$, else at least one of the terms of the arithmetic progression will be divisible by q.)

2.9. Magic squares of primes. A *magic square* is an n-by-n array of distinct integers such that the sum of the numbers in any row, or in any column, or in either diagonal, equals the same constant. These have been very popular in the recreational mathematics literature (see, e.g., [**7**]). Here are two small examples.

17	89	71
113	59	5
47	29	101

41	89	83
113	71	29
59	53	101

Examples of 3*-by-*3 *magic squares of primes.*

Do you recognize the primes involved? Do you notice any similarities with the smallest 3-by-3 GAPs of primes (see Section 2.2)? The reason is that every 3-by-3 GAP can be rearranged to form a 3-by-3 magic square and vice-versa!

37	83	97	41
53	61	71	73
89	67	59	43
79	47	31	101

41	71	103	61
97	79	47	53
37	67	83	89
101	59	43	73

Examples of 4*-by-*4 *magic squares of primes.*

It has long been known that there are n-by-n magic squares for any $n \geq 3$. If the entries are $m_{i,j}$, $1 \leq i, j \leq n$, then the square with (i, j)th entry $a + m_{i,j}b$ is also an n-by-n magic square. The Green-Tao theorem implies that there are infinitely many pairs of integers a, b for which all of the integers $a + \ell b$, $\min_{i,j} m_{i,j} \leq \ell \leq \max_{i,j} m_{i,j}$ are prime, and this yields infinitely many n-by-n magic squares of primes.

By the obvious modifications of this argument we can show that if there is a magic cube of a given size then there are infinitely many magic cubes of primes of the same size, and the same is true for higher dimensional objects of this type. *Bi-magic* squares are magic squares for which the squares of the entries also form a magic square. Since $\sum(a + m_{i,j}b)^2 = a^2 \sum 1 + 2ab \sum m_{i,j} + b^2 \sum m_{i,j}^2$ (summing over any appropriate domain), we see that if there exist n-by-n bi-magic squares then there are infinitely many n-by-n bi-magic squares of primes.

The examples of 4-by-4 magic squares of primes given above have another property. The primes involved are *all of the primes between* 31 *and* 101, and *between* 37 *and* 103, respectively. Proving that there are infinitely many examples of n-by-n magic squares in which the primes involved are all of the primes from some interval does seem, to me, to be beyond the scope of applications of the Green-Tao theorem.

2.10. Other prime patterns? I am sure that there are other beguiling consequences of the Green-Tao theorem, and I hope that the reader will find some!

There are prime patterns of this type which do not *seem* to follow from the Green-Tao theorem but do follow easily from older work (as Antal Balog remarked to me): from [**2**], though seemingly not from [**4**], one can immediately deduce that there are infinitely many sets of four primes such that the sum of any three is also prime.

3. IN WHAT SITUATIONS DO WE EXPECT TO HAVE PRIME PATTERNS? THE PRIME *k*-TUPLETS CONJECTURE AND BEYOND. To generalize the notion of an arithmetic progression $a + jd, 0 \le j \le k-1$, which is a set of k linear polynomials in $\mathbb{Z}[a, d]$, we consider the k-tuple of linear polynomials $L_1(a_1, \dots, a_n), \dots, L_k(a_1, \dots, a_n) \in \mathbb{Z}[a_1, \dots, a_n]$. We wish to determine whether there are infinitely many sets of integers $\{a_1, a_2, \dots, a_n\}$ for which each $|L_j(a_1, a_2, \dots, a_n)|$ is prime. There are examples for which there are only a finite number of sets $\{a_1, a_2, \dots, a_n\}$ with each $|L_j(a_1, a_2, \dots, a_n)|$ prime; for example, if we have the polynomials $L_j(a) = dj + a$ for $1 \le j \le p$, where prime p does not divide integer d, then p always divides the value of one of the linear forms no matter what the choice of a. To exclude this possibility we call the set of linear forms *admissible* if, for all primes p, there are integers $a_1, \dots, a_n$ such that $\prod_{j=1}^{k} L_j(a_1, \dots, a_n)$ is not divisible by p. The *extended prime k-tuplets conjecture* states that if the set of linear forms is admissible then there are infinitely many choices of integers $a_1, \dots, a_n$ for which each $|L_j(a_1, \dots, a_n)|$ is prime.

In this subsection we will prove various results concerning extensions of the prime patterns discussed in Section 3. (The subsection numbers are coordinated so as to refer to the same problem in each section; for example Sections 2.3 and 3.3 both discuss Balog cubes).

3.1. All arithmetic progressions of primes are finite. Suppose that all the terms of the arithmetic progression $q + jd$, $0 \le j \le k-1$, with $d \ge 1$, are prime. In Section 2.8 we noted that q divides $q + qd$, so $k \le q$. We will now improve this bound by showing that $k \le p$, where p is the smallest prime that does not divide d. Suppose that $k \ge p$ so that $p \le k \le q$, and therefore $p < q + jd$ for all $j \ge 1$. Now, for any prime ℓ that does not divide d, one in every ℓ consecutive terms of the arithmetic progression $q + jd$ is divisible by ℓ, so, in particular, some prime $q + jd$, $0 \le j \le p-1$ ($\le k-1$), is divisible by p, and thus must be p. Therefore $q = p$ (since we noted that the other terms of the arithmetic progression are $> p$) and our claim follows.

Now we show that an arithmetic progression in which the absolute values of the terms are all prime has no more than $2p-1$ terms, where p is the smallest prime that does not divide d. If it had $\ge 2p$ terms then two would be divisible by p, and so would have to be $-p$ and p, implying that d divides $2p$. This is impossible, for if $d = 1$ or 2 then the arithmetic progression would contain the non-prime 1, and if $d = p$ or $2p$ it would contain 0, $-3p$ or $3p$. The upper bound $2p-1$ can be attained: by $-3, 2, 7$ for $p = 2$, by $-13, -5, 3, 11, 19$ for $p = 3$, and by $5 + 12j$, $-4 \le j \le 4$ for $p = 5$. For arbitrary p let m be the product of the primes $< p$. The set

$$\{p + jmt,\ -(p-1) \le j \le p-1\}$$

of linear forms in $\mathbb{Z}[t]$ is admissible (as may be seen by selecting $t = 1$ so that the product of the linear forms is not divisible by p, and $t = 0$ to test for non-divisibility by any other prime). The extended prime k-tuplets conjecture thus predicts that there are infinitely many choices of integers n such that $|p + jmn|$ is prime for all j in the range $-(p-1) \le j \le p-1$.

3.2. All GAPs of primes are finite. A GAP is a set of integers of the form $\{a + n_1 b_1 + \cdots + n_d b_d :\ 0 \le n_1 \le N_1 - 1, \dots, 0 \le n_d \le N_d - 1\}$ for given integers $a, b_1, b_2, \dots, b_d$, and integers $N_1, N_2, \dots, N_d \ge 2$. By the argument of Section 3.1 we know that if the terms of the GAP are distinct primes then $N_i \le p_i$ for each $i = 1, 2, \dots, d$, where p_i is the least prime that does not divide b_i. If $d \ge 2$ then we can improve this

to $N_i \le p_i - 1$: for if $N_i = p_i$ then, for any fixed $n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_d$, we have an arithmetic progression of length p_i and common difference b_i (as n_i varies), so that some element of that arithmetic progression is divisible by p_i and thus must be p_i, contradicting the hypothesis that the elements of the GAP are distinct primes.

Using a tool of additive combinatorics, we will show that if the terms of our GAP are distinct primes and

$$\sum_{i:\ p \text{ does not divide } b_i} (N_i - 1) \ge p - 1, \tag{3.1}$$

then some element of our GAP must equal p. Moreover we will show that if we have a GAP of distinct primes then

$$\sum_{i:\ p \text{ does not divide } b_i} (N_i - 1) \le p - 1; \tag{3.2}$$

and if (3.1) holds then p does not divide $b_1 b_2 \dots b_d$. The set

$$\{p + n_1 b_1 + n_2 b_2 :\ 0 \le n_1 \le N_1 - 1,\ 0 \le n_2 \le N_2 - 1, \{n_1, n_2\} \ne \{0, 0\}\} \subseteq \mathbb{Z}[b_1, b_2]$$

where $(N_1 - 1) + (N_2 - 1) = p - 1$ is admissible (for example, by taking $b_1 \equiv b_2 \equiv 1 \pmod p$, and $b_1 \equiv b_2 \equiv 0 \pmod q$ whenever $q \ne p$), and so we believe that there are infinitely many GAPs of distinct primes of this form, by the extended prime k-tuplets conjecture (and thus (3.1) is best possible).

Now, let I_p be the set of $i \in \{1, \dots, d\}$ for which p does not divide b_i. The sets $\{n_i b_i \pmod p :\ 0 \le n_i \le N_i - 1\}$ each consist of $\min\{N_i, p\}$ distinct elements. We define the addition of two sets A and B of residues mod p by

$$A + B = \{a + b \pmod p :\ a \in A, b \in B\}.$$

The Cauchy-Davenport theorem tells us that if A, B are sets of residues mod p then $|A + B| \ge \min\{|A| + |B| - 1, p\}$. Since our GAP $(\text{mod } p)$ equals

$$\{a\} + \{n_1 b_1 \pmod p :\ 0 \le n_1 \le N_1 - 1\} + \cdots + \{n_d b_d \pmod p :\ 0 \le n_d \le N_d - 1\},$$

we deduce, by induction, that the size of our GAP $(\text{mod } p)$ is

$$\ge \min\{p, 1 + \sum_{i \in I_p} (N_i - 1)\}.$$

In particular it contains 0 $(\text{mod } p)$ once $\sum_{i \in I_p} (N_i - 1) \ge p - 1$, that is, (3.1) holds, and so the GAP contains p.

We have just seen that if (3.1) holds then some element $a + m_1 b_1 + \cdots + m_d b_d$ of our GAP is divisible by p; if p divides b_j then $a + m_1 b_1 + \cdots + m_{j-1} b_{j-1} + n_j b_j + m_{j+1} b_{j+1} + \cdots + m_d b_d$ is divisible by p for each $0 \le n_j \le N_j - 1$, contradicting the fact that the GAP consists of distinct primes.

We have now seen that if $\sum_{i \in I_p} (N_i - 1) \ge p$ then there is some element $a + m_1 b_1 + \cdots + m_d b_d$ which is divisible by p and $I_p = \{1, 2, \dots, d\}$. Adding the sets

$$\{a\} + \{n_1 b_1 \pmod p :\ 0 \le n_1 \le N_1 - 1\} + \dots$$
$$+ \{n_{d-1} b_{d-1} \pmod p :\ 0 \le n_{d-1} \le N_{d-1} - 1\}$$
$$+ \{n_d b_d \pmod p :\ 0 \le n_d \le N_d - 1,\ n_d \ne m_d\},$$

we find, by the Cauchy-Davenport theorem, that there is a second element of our GAP that is divisible by p, contradicting the fact that the GAP consists of distinct primes.

3.3. All Balog cubes are finite. Were there an infinite Balog cube of primes it would contain an infinite arithmetic progression of primes, which is impossible (see Section 3.1).

3.4. Infinite sets of primes, averaging in pairs. We will construct an infinite sequence of primes $p_1 = 3 < p_2 = 7 < p_3 = 19 < p_4 < \cdots$ such that all $p_{i,j} = (p_i + p_j)/2$ are prime and distinct for $1 \le i \le j$ using the extended prime k-tuplets conjecture. We find p_k for each $k \ge 4$ by induction: Let m_k be the product of the primes $\le k$ and let n_k be the product of the primes in m_k of the form $p_{1,j}$ for $1 \le j \le k-1$. Let a_k be the least residue mod $4m_k$ for which

$$a_k \equiv \begin{cases} 3 & (\text{mod } 4m_k/n_k) \\ 7 & (\text{mod } n_k/(5, n_k)) \\ 19 & (\text{mod } (5, n_k)) \end{cases}$$

We look for the desired p_k to be of the form $a_k + 4rm_k$ for some integer r, so we need to find r such that the linear forms $a_k + 4rm_k$ and $(a_k + p_j)/2 + 2rm_k$ for $1 \le j \le k-1$ are all prime; and so we now show that this set is admissible. If $p > k$ then p does not divide $2m_k$ and thus we can find an integer r for which none of the linear forms is divisible by p, since there are only $k < p$ forms in our set. If p divides m_k/n_k then $a_k + 4rm_k \equiv 3 \not\equiv 0 \pmod p$ and $(a_k + p_j)/2 + 2rm_k \equiv p_{1,j} \not\equiv 0 \pmod p$. Similarly if p divides $n_k/(n_k, 5)$ then $a_k + 4rm_k \equiv 7 \not\equiv 0 \pmod p$ and $(a_k + p_j)/2 + 2rm_k \equiv p_{2,j} \not\equiv 0 \pmod p$. Finally if 5 divides n_k then $a_k + 4rm_k \equiv 19 \not\equiv 0 \pmod 5$ and $(a_k + p_j)/2 + 2rm_k \equiv p_{3,j} \not\equiv 0 \pmod 5$. Hence such a p_k exists by the extended prime k-tuplets conjecture.

3.5. Averaging sets of integers must be finite. We now show that averaging sets of integers, and thus averaging sets of primes, must be finite. Suppose that $a < b$ are two elements of A and let p be the smallest prime which does not divide $b - a$. If $|A| \ge p + 1$ then we should have that the average of every p elements of A is an integer. So let S be a subset of $A \setminus \{a, b\}$ with $p - 1$ elements. Then p divides $a + \sum_{s \in S} s$ and $b + \sum_{s \in S} s$, so that p divides their difference, a contradiction. Therefore $|A| \le p$.

3.6a. Only a finite number of initial values of a polynomial can be prime. For $f(x) \in \mathbb{Z}[x]$ of degree $d \ge 1$, suppose that $|f(0)|, |f(1)|, \dots, |f(k)|$ are distinct primes, and let $p = f(0)$. Then $f(p) \equiv f(0) \equiv 0 \pmod p$ and so $|f(p)|$ cannot be a prime distinct from $p = |f(0)|$. Therefore $k \le p - 1$.

3.6b. An arbitrary number of initial values of a monic degree *d* polynomial can be prime. Let $g(x) \in \mathbb{Z}[x]$ be any monic polynomial which, for each prime p, is NOT a permutation polynomial for the residues mod p (g is a *permutation polynomial* if g permutes the residues mod p or, equivalently, if the map $g : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ is

1-to-1). The easiest way to achieve this is to take $g(x) = x(x-1)h(x)$ for some monic polynomial $h(x)$ of degree $d-2$, since then $g(0) = g(1) = 0$ and so the map cannot be 1-to-1. The point of the requirement is that for every prime p there exists m_p such that $g(r) \not\equiv m_p \pmod p$ for all integers r, and so, for any integer k, the set of linear forms $a + g(0), a + g(1), \dots, a + g(k)$ is admissible. Therefore the prime k-tuplets conjecture predicts that there will be infinitely many prime k-tuplets of this form. Take such an integer a and let $f(x) = g(x) + a$ to obtain a monic polynomial of degree d for which $f(0), f(1), \dots, f(k)$ are all prime.

A similar argument indicates that the bound in Section 3.6a is best possible: For any given prime p let $g(x)$ be any monic polynomial with $g(0) = p$, and let $k = p-1$. For any prime q, the set $\{g(j)/j \pmod q : 1 \le j \le \min\{q-1, k\}\}$ contains $< q$ elements and so there exists a_q such that $-a_q \pmod q$ is not in this set. Therefore if $1 \le j \le k$ and q does not divide j then $g(j) + ja_q \not\equiv 0 \pmod q$. If q divides j then $q \ne p$, and therefore $g(j) + ja_q \equiv g(0) = p \not\equiv 0 \pmod q$. Therefore the set $g(1) + a$, $g(2) + 2a, \dots, g(k) + ka$ of linear forms is admissible, and so we expect infinitely many integers a for which they are all prime. Take such an integer a and let $f(x) = g(x) + ax$ to obtain a monic polynomial of degree d for which $f(0), f(1), \dots, f(k)$ are all prime with $k = |f(0)| - 1$.

4. THE NUMBER OF SUCH PRIME PATTERNS, AND THE SMALLEST SUCH PATTERN. Given an admissible k-tuple of linear polynomials

$$L_1(a_1, \dots, a_n), \dots, L_k(a_1, \dots, a_n) \in \mathbb{Z}[a_1, \dots, a_n],$$

with $k > n$, we wish to determine the number of sets of integers $\{a_1, \dots, a_n\} \in \mathcal{A}$ for which each $|L_j(a_1, \dots, a_n)|$ is prime. Here $\mathcal{A}$ is some "reasonable" domain (for example, convex) such that, for almost all $(a_1, \dots, a_n) \in \mathcal{A}$, each $|L_j(a_1, \dots, a_n)|$ is about size x.

The prime number theorem states that there are about $x/\log x$ primes up to x; in other words, 1 in every $\log x$ of the integers close to x is prime. Thus, if the values of the $L_j(\mathbf{a})$, $\mathbf{a} \in \mathcal{A}$, could be considered independently then we would expect that 1 in every $(\log x)^k$ such k-tuples is prime. However in many examples such forms are obviously dependent; for example, the pair of linear forms $n, n+2$ are coprime to 3 for 1 in every 3 choices of n, whereas two independent linear forms are coprime to 3 for 4 out of every 9 choices (note $4/9 = (2/3)^2$). We can similarly adjust our probabilities for divisibility by each prime, and end up with the conjecture:

The proportion of $\{a_1, \dots, a_n\} \in \mathcal{A}$ for which each $|L_j(a_1, \dots, a_n)|$ is prime is about

$$\frac{1}{(\log x)^k} \prod_{p \text{ prime}} \frac{\text{Prob}((L_1(\mathbf{a}) \dots L_k(\mathbf{a}), p) = 1 : 0 \le a_1, \dots, a_n \le p-1)}{\text{Prob}((\ell, p) = 1 : 0 \le \ell \le p-1)^k}. \tag{4.1}$$

(Here the notation Prob(Event E : Range R) means the probability that event E is true if the variables are chosen from the range R, with each assignment of values of the variables from R having equal probability.)

In order to use this to make predictions (as at the conclusion of each subsection in Section 2) we need to give reasonably good approximations to each term in the formula. We start by evaluating the easier terms, and by clarifying how to determine some other terms.

We now take $\mathcal{A}$ to be the set of integers $\{a_1, \dots, a_n\}$ for which each

$$|L_j(a_1, \dots, a_n)| \le x.$$

It is not difficult to show that $|\mathcal{A}| \sim c_L x^n$, for some constant $c_L > 0$.

For a given prime p we have $\text{Prob}((\ell, p) = 1 : 0 \le \ell \le p-1) = 1 - 1/p$. Defining $\omega(p)$ to be the number of choices of $a_1, \dots, a_n \pmod p$ for which $L_1(\mathbf{a}) \dots L_k(\mathbf{a})$ is divisible by p, we have $\text{Prob}((L_1(\mathbf{a}) \dots L_k(\mathbf{a}), p) = 1 : 0 \le a_1, \dots, a_n \le p-1) = 1 - \omega(p)/p^n$. Using these remarks we can rewrite (4.1) as:

The number of $\{a_1, \dots, a_n\} \in \mathcal{A}$ for which each $|L_j(a_1, \dots, a_n)|$ is prime is about

$$c_L \prod_{p \text{ prime}} \left\{ \left(1 - \frac{\omega(p)}{p^n}\right) \left(1 - \frac{1}{p}\right)^{-k} \right\} \frac{x^n}{(\log x)^k} \tag{4.2}$$

This is the *quantitative prime k-tuplets conjecture*.

We need two estimates concerning primes. The prime number theorem states that $\pi(y)$, the number of primes up to y, is about $y/\log y$. Hence most primes up to y will have size near y (by which I mean larger than $y/\log y$), so we can deduce that the product of the primes up to y is about $y^{y/\log y} = e^y$. Mertens' theorem, which can also be deduced from the prime number theorem (though preceeds it historically), states that

$$\prod_{p \le y} \left(1 - \frac{1}{p}\right)^{-1} \approx e^\gamma \log y.$$

Armed with these two estimates let us proceed to approximate (4.2) for the simplest example discussed in this article:

4.1. Arithmetic progressions of primes. There are $\sim x^2/(2(k-1))$ pairs of integers $a, d \ge 1$ with $a + (k-1)d \le x$, and therefore we predict that the number of k-term arithmetic progressions of distinct primes all $\le x$ is given by (4.2) with $c_L = 1/(2(k-1))$ and $n = 2$.

Next we wish to evaluate $\omega(p)$. For each prime $p \le k$ we have seen that at least one of $a + jd, 0 \le j \le p-1 \le k-1$, is divisible by p unless p divides d. In that case one of them is divisible by p if and only if p divides a. Therefore $\omega(p) = p(p-1) + 1$, and so

$$\prod_{p \text{ prime}, \ p \le k} \left\{ \left(1 - \frac{\omega(p)}{p^2}\right) \left(1 - \frac{1}{p}\right)^{-k} \right\} = \prod_{p \text{ prime}, \ p \le k} \frac{1}{p} \left(1 - \frac{1}{p}\right)^{-(k-1)}, \tag{4.3}$$

which is roughly of size $e^{-k} \cdot (e^\gamma \log k)^k = (e^{\gamma-1} \log k)^k$ by the two estimates for prime numbers given above.

For the primes $p > k$ we note that p divides some $a + jd$, $0 \le j \le k-1$, if and only $a \equiv -jd \pmod p$ for some $0 \le j \le k-1$. When p does not divide d these k values mod p are distinct so we deduce that $\omega(p) = k(p-1) + 1$, and so the pth term in (4.2), for $p > k$, equals

$$\left(1-\frac{k-1}{p}\right)\left(1-\frac{1}{p}\right)^{-(k-1)}. \tag{4.4}$$

Taking logarithms and using the Taylor series for $\log(1-x)$, we find that the first term of each Taylor series, $\frac{k-1}{p}$, cancels, and that what remains is dominated by the $\frac{k^2}{p^2}$ term. We can now bound the total contribution of ALL the primes $p > k$ to (4.2): The logarithm of their contribution is bounded by a constant multiple of $\sum_{p>k} \frac{k^2}{p^2}$; and the prime number theorem can be used to show that $\sum_{p>k} \frac{1}{p^2}$ is bounded by a constant multiple of $\frac{1}{k \log k}$.

Collecting together all these evaluations, we have shown that our prediction for the number of k-term arithmetic progressions of distinct primes all $\le x$ is about

$$\left(e^{\gamma-1}\frac{\log k}{\log x}\right)^k x^2. \tag{4.5}$$

We have avoided many technicalities by the use of the word "about"; for example, the contributions of c_L, and of the primes $p > k$, are negligible at this level of precision. To be more precise we could replace '$e^{\gamma-1}$' by '$e^{\gamma-1} + o(1)$', where $o(1)$ represents some function which tends to 0 as k and x get large.

If we take $x = ((e^{1-\gamma}/2 - \epsilon)k)^{k/2}$ then (4.5) gives roughly $e^{-c\epsilon k}$ such k-term arithmetic progressions of primes, for some constant $c > 0$. Note that $e^{-c\epsilon k}$ is very small, far less than 1, so our prediction is that there should be no such arithmetic progressions up to this x, or at most very few. On the other hand if we take $x = ((e^{1-\gamma}/2 + \epsilon)k)^{k/2}$ then we expect more than $(1 + c\epsilon)^k$ such k-term arithmetic progressions of primes. Now $(1 + c\epsilon)^k$ is exponential in k, so once k is large this is very very large, hence we are pretty sure that at least one such k-term arithmetic progressions of primes exists. This sharp transition around $((e^{1-\gamma}/2)k)^{k/2}$ (the value given in (2.1)) is very typical in these predictions and the evidence from our data in Section 2.1 gives us confidence that we are guessing correctly.

We end this subsection by being more precise than in (2.1): we conjecture that the smallest k-term arithmetic progression of distinct primes will have largest prime $((e^{1-\gamma}/2 + o(1))k)^{k/2}$.

4.2. The other patterns. In all of the other cases we can understand most of the terms in (4.2) in an analogous way: The terms in the product for the primes $p > k$ similarly contribute little. The product of $(1 - 1/p)^{-k}$ over the primes $p \le k$ can again be understood by Mertens' theorem. We can even determine the total contribution of the "small" primes p in some generality:

The set of linear forms is admissible if and only if $\omega(p) < p^n$ for every prime p. Since $\omega(p)$ is an integer this implies that $\omega(p) \le p^n - 1$ and so $1 - \omega(p)/p^n \ge 1/p^n$. This implies that, for fixed $\epsilon > 0$,

$$1 \ge \prod_{\substack{p<\epsilon k/n \\ p \text{ prime}}} \left(1-\frac{\omega(p)}{p^n}\right) \ge \prod_{\substack{p<\epsilon k/n \\ p \text{ prime}}} \frac{1}{p^n} \approx \frac{1}{e^{2\epsilon k}},$$

and thus this product is not large enough to be significant in the end result, if ϵ is sufficiently small.

It therefore only remains to estimate

$$\text{Prob}((L_1(\mathbf{a}) \dots L_k(\mathbf{a}), p) = 1 : \ 0 \le a_1, \dots, a_n \le p-1)$$

in the range $\epsilon k/n < p < k$.

Each prime pattern requires different, and nontrivial, combinatorial arguments to estimate these probabilities, and a full explanation is perhaps beyond the scope of this article. However I find the result in Section 2.6, on consecutive prime values of polynomials, to be the most interesting, so let us work out this one final case.

4.3. Initial polynomial values. We write our polynomial as

$$f(X) = X^d + \sum_{i=0}^{d-1} a_i X^i$$

with $d \geq 2$, and assume that each a_i is a nonnegative integer, so that $n = d$. To ensure that $f(0), f(1), \dots, f(k-1) \leq x$ we want $(k-1)^d + \sum_{i=0}^{d-1} a_i (k-1)^i \leq x$ (so we take x substantially larger than $(k-1)^d$); the number of such sets of integers $a_0, a_1, \dots, a_d \geq 1$ is about

$$\frac{1}{d!} \prod_{i=0}^{d-1} \frac{x}{(k-1)^i} = \frac{x^d}{d!\,(k-1)^{d(d-1)/2}},$$

so that $c_L = 1/(n!\,(k-1)^{n(n-1)/2})$, which is negligible if we take $k \geq 4(d \log d)^2$.

We now determine

$$\text{Prob}((f(0)f(1)\dots f(k-1), p) = 1 : \ 0 \leq a_0, a_1, \dots, a_{d-1} \leq p-1).$$

Since $f(r+p) \equiv f(r) \pmod p$ for each r, this is equal to

$$\text{Prob}((f(0)f(1)\cdots f(\ell-1), p) = 1 : \ 0 \leq a_0, a_1, \dots, a_{d-1} \leq p-1),$$

where $\ell = \min\{k, p\}$. Define $g(X) := \sum_{j=0}^{p-2} b_j X^j$ where $b_j = \sum_{i:\ i \equiv j \pmod{p-1}} a_i$, so that $f(n) \equiv g(n) \pmod p$ whenever $1 \leq n \leq p-1$, since $n^i \equiv n^j \pmod p$ for all $0 \leq j \leq p-2$ and $i \equiv j \pmod{p-1}$. As $f(0) \equiv a_0 \pmod p$, we therefore wish to determine

$$\text{Prob}((a_0 g(1)\cdots g(\ell-1), p) = 1 : \ 0 \leq a_0, a_1, \dots, a_{d-1} \leq p-1)$$

For $0 \leq j \leq J := \min\{p-2, d-1\}$ the value of $b_j \pmod p$ runs equally often through each value $\pmod p$ as the $a_i, i \equiv j \pmod p$ do: to see this, simply fix $a_i, i \equiv j \pmod{p-1}, i > j$, and notice that $b_j \pmod p$ runs through each value $\pmod p$ as a_j does. Now if $d < p$ then $g(0) = b_0 = a_0$ and so we want

$$\text{Prob}((g(0)g(1)\dots g(\ell-1), p) = 1 : \ 1 \leq b_0, b_1, \dots, b_J \leq p).$$

If $d \geq p$ then $b_0 \pmod p$ runs equally often through each number $\pmod p$ as the a_i, $i \equiv 0 \pmod{p-1}, i > 0$ do, independent of a_0 (to see this, modify the above proof but this time fix all the $a_i, i \neq p-1$) and so we want

$$\text{Prob}((g(1)\cdots g(\ell-1), p) = 1 : \ 1 \leq b_0, \dots, b_J \leq p)$$
$$\times \text{Prob}((a_0, p) = 1 : \ 1 \leq a_0 \leq p).$$

The matrix that determines the values $g(0), g(1), \dots, g(\ell-1)$ from $b_0, b_1, \dots, b_J$ is a Vandermonde matrix and therefore of full rank. Hence, as we run through the set of all possible values of $(b_0, b_1, \dots, b_J) \pmod p$ we run through all possible values

of $(g(0), g(1), \dots, g(\ell-1))$ (mod p) each exactly $p^{J-\ell+1}$ times. Therefore our probability equals $(1-1/p)^{\ell}$ (in both cases). This implies that the product over primes in (4.2) is exactly $\prod_{p<k}\left(1-\frac{1}{p}\right)^{-(k-p)}$, which can be evaluated using Mertens' theorem and the prime number theorem.

Thus (4.2) is about $(\lambda e^{\gamma})^k$ when $x = (\lambda k/n)^{k/n}$, and so we deduce (2.6) with d replaced by n.

5. CONCLUDING REMARKS. The purpose of this article has been to exhibit how the amazing result of Green and Tao, that there are infinitely many k-term arithmetic progressions of primes, yields some entertaining consequences, allowing us to prove that there are primes in all sorts of mathematically and aesthetically desirable patterns. But their work has not stopped with [**4**]. Indeed Green and Tao are audaciously planning an assault on arguably the greatest of all conjectures about the primes, the extended prime k-tuplets conjecture. They have, of course, already succeeded when the set of linear forms is $a + jd$, $0 \le j \le k-1$, but, in mid-2006, they released a preprint [**6**] in which they describe a plausible program for going much further. There are certain cases that they are unable to attack as yet: that there are infinitely many pairs of primes p, $p+2$ (*the twin prime conjecture*); for any large even integer N there are pairs of primes p, $N-p$ (*the Goldbach conjecture*), and that there are infinitely many pairs of primes p, $2p+1$ (*Sophie Germain twins*). Note that these are all examples of *difficult pairs* of linear forms, L_1, L_2, in that there exist nonzero integers a, b, c for which $aL_1 + bL_2 = c$. Green and Tao now believe that they will prove the extended prime k-tuplets conjecture for any admissible k-tuple of linear forms that does not contain a difficult pair.

In their preprint [**5**], Green and Tao take a first giant step towards their ambitious program, proving results that go well beyond [**4**]. Peter Sarnak has informed me of several delightful applications of the results in [**5**]:

5.1. Pythagorean triples. It is well known that any solution to $x^2 + y^2 = z^2$ in coprime integers must be of the form

$$x = r^2 - s^2, \quad y = 2rs, \quad z = r^2 + s^2,$$

where r and s are coprime integers with $r+s$ odd. The area of the right-angled triangle with sides x, y, and z is given by

$$A = \frac{xy}{2} = rs(r+s)(r-s), \tag{5.1}$$

and must be divisible by 6 since one of r and s must be even (as $r+s$ is odd), and since one of $r, s, r^2 - s^2$ must be divisible by 3. Hence we can ask how few prime factors can $A/6$ have? In (5.1) we saw that A is the product of four factors which are linear polynomials in r and s, so there can be only finitely many pairs r, s for which $A/6$ has fewer than three prime factors. Calculations reveal that $A/6 = 1$ only for the $(3, 4, 5)$ triangle, and that $A/6$ has exactly one prime factor only for the $(5, 12, 13)$ triangle. The only Pythagorean triples for which $A/6$ has exactly two prime factors are

$$(8, 15, 17),\ (7, 24, 25),\ (12, 35, 37),\ (20, 21, 29),\ (11, 60, 61),\ \text{and } (13, 84, 85).$$

We believe that there are infinitely many Pythagorean triples for which $A/6$ has exactly three prime factors, since the prime k-tuplets conjecture predicts that there are infinitely many prime triplets $p-6$, p, $p+6$ and, when we do have such a triplet,

we can take $r = p$ and $s = 6$ above. Unfortunately we still cannot prove that there are infinitely many such prime triplets.

What can now be proved unconditionally is that there are infinitely many Pythagorean triples such that $A/6$ has exactly four prime factors; for instance when $r = 2p$ and $s = 3q$ where $p, q, 2p + 3q$, and $2p - 3q$ are all prime. A consequence of [**5**] is that there are indeed infinitely many such prime quadruplets. (This all appeared in Ben Tsou's junior undergraduate thesis at Princeton.)

5.2. Matrices of a given determinant. The determinant of an n-by-n matrix of odd integers must be divisible by 2^{n-1}. Nevo and Sarnak [**8**] have shown that for any integer m divisible by 2^{n-1}, there are infinitely many n-by-n matrices of determinant m whose entries are distinct (odd) primes, for any $n \geq 3$.

If Green and Tao can indeed prove the extended prime k-tuplets conjecture for any admissible k-tuple of linear forms that does not contain a difficult pair, then this will be an even more widely applicable theorem than any of their remarkable results to date (for example, from this one can deduce that there are infinitely many monic polynomials of degree d for which the first k values give distinct primes). Green and Tao even believe that they will be able to prove that (4.2) is an accurate approximation for the number of such prime k-tuples. This is another leap forward far beyond the horizon, and will no doubt give rise to many other extraordinary patterns of primes.

ACKNOWLEDGEMENTS. Many thanks to Antal Balog, Ben Green, Peter Sarnak, and Terry Tao for their comments on these notes and for their ideas which were subsequently incorporated; and to Tomás Oliveira e Silva and Tony Noe for their help with the data for Sections 2.4 and 2.5. I computed many of the examples in the article myself, but I also found several at the websites [**1**], [**3**], [**7**], [**9**] which are marvellous, widely available resources. I did not give credit to the discoverers of each example since that would arguably have disturbed the flow of the article. However, please go to those websites for appropriate credits and lots more examples.

REFERENCES

1. J. K. Andersen, Primes in Arithmetic Progression Records, available at `http://hjem.get2net.dk/jka/math/aprecords.htm`.
2. A. Balog, The prime k-tuplets conjecture on average, in *Analytic Number Theory*, B. C. Berndt et al., eds., Birkhäuser, Boston, 1990, 165–204.
3. C. Caldwell, The prime pages, available at `http://primes.utm.edu/`.
4. B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Ann. Math.* (to appear); also available at `http://xxx.arxiv.org/math.NT/0404188`.
5. ———, Quadratic Uniformity of the Mobius Function, available at `http://xxx.arxiv.org/math.NT/0606087`.
6. ———, Linear Equations in Primes, available at `http://xxx.arxiv.org/math.NT/0606088`.
7. H. Heinz, Magic Squares, Magic Stars and Other Patterns, available at `http://www.geocities.com/~harveyh/`.
8. A. Nevo and P. Sarnak, Prime Matrices, (to appear).
9. P. Ribenboim, *The Little Book of Bigger Primes*, 2nd ed., Springer, New York, 2004.
10. A. Zimmerman, Prime Generating Polynomials programming competition (03/13/2006–06/13/2006), Al Zimmerman's Programming Contests, available at `http://www.recmath.org/contest/index.php`.

ANDREW GRANVILLE received his B.A. from Cambridge University in 1983, and his Ph.D. from Queen's University in Canada in 1987. He is now the Canadian Research Chair in number theory at the Université de Montréal. He was seduced into studying Fermat's Last Theorem for his doctorate by the engaging writing of his Ph.D. advisor, Paulo Ribenboim, and shares with him a desire to understand and explain the distribution of prime numbers.

Département de Mathématiques et Statistique, Université de Montréal,
CP 6128 succ Centre-Ville, Montréal, QC H3C 3J7, Canada
andrew@dms.umontreal.ca