WHAT IS... the Parity Phenomenon?

John Friedlander and Henryk Iwaniec

Sometimes it almost seems as though there is a ghost in the House of Prime Numbers.

When you enter the front door of the house you come first to the Elementary Room. This is the oldest part of a structure which has been assembled piecemeal over a period of many centuries. In the Elementary Room lots of people just sit around counting prime numbers, and one of their favourite ways to do so is with sieve methods. This circle of ideas dates from the inclusion-exclusion principle of Eratosthenes. Perhaps it is revered simply because it is about as old as the house itself but, whatever the reason, people have been ignoring their failure to make things work as hoped and continue to push forward their efforts. Only with Brun, a little less than one hundred years ago, did one begin to have some movement toward success.

Recall that the sieve of Eratosthenes is applied by starting with your favourite sequence of integers, then casting out the multiples of each small prime, noting that some integers have been cast out more than once and then using inclusion-exclusion to rectify the count so that each unwanted number is cast out exactly once. One can describe the result of this both precisely and rather succinctly by using the Möbius function. However, the result is worthless for purposes of estimation because it expresses the number of primes in the sequence as a sum having a huge number of terms not so easy to add, and such that even an excellent approximation to the size of each individual term is not enough to give a useful approximation to the sum.

John Friedlander is University Professor of Mathematics at the University of Toronto. His email address is `frdlndr@utsc.utoronto.ca`.

Henryk Iwaniec is State of New Jersey Professor of Mathematics at Rutgers University. His email address is `iwaniec@math.rutgers.edu`.

Brun got the idea of replacing the Möbius function by another arithmetic function (the sieve weights) having somewhat similar properties but supported only on a finite set of integers so that he could truncate the sum at will and get a manageable error. It was too much to hope (indeed not possible) to do this in such a way that one could get an identity for the number of primes, as one did with the Möbius function, but it turned out that there were lots of ways to choose the sequence of sieve weights so that one got an upper bound (or, choosing differently, a lower bound) for the number of survivors of the sifting process.

During the next few decades the main emphasis in sieve theory was on the search for optimal weights, those that would make the upper bound as small as possible and others that would make the lower bound as large as possible. Because these weights are somewhat complicated, it is not easy to see at once whether or not a specific lower bound might turn out to be negative, in which case the bound is even worse than trivial. The upper bound weights do much better, at least in some sense. In practice, even an upper bound very far from the truth will at least give a result of the right order of magnitude. After the introduction of the Selberg sieve one could begin in some important problems to approach the optimal sieve weights.

There is a second aspect to the sieve problem. Just as in the original inclusion-exclusion procedure, the new sieve weights reduce the problem of counting primes to the problem of counting multiples of various integers d, those composed of the small primes. As one might expect, the larger you can take d, say all $d < D$, and still get useful estimates for the number of its multiples, the more successful will be the sieving procedure. The best one can do in this regard depends on the sequence of integers with which you began, and this largest value of D is called the level of distribution of the sequence. There is a natural limitation to how large the level can be. To take a simple example, suppose you consider a segment of 409 consecutive integers and ask how many multiples of 666 it contains. The answer is either zero or one and depends on which segment of integers you started with. But the sieve wants to assume that the number of multiples is 409/666.

Once one is armed with optimal sieve weights and given a sequence with optimal level of distribution, one naturally asks:

1) How large is the upper bound one obtains for the number of primes?

2) Do we get a positive lower bound for the number of primes?

The answers are:

1) Just twice what one expects (from heuristic arguments based on randomness assumptions).

2) No, but we miss by the narrowest of margins.

To see that these two answers cannot be improved one may consider the counterexamples provided, in the first case, by the sequence of positive integers having an odd number of prime factors and, in the second case, by those with an even number of prime factors.

These observations are due to Selberg (around 1949), who named the phenomenon the "principle of parity" (the name came quite a bit later; see Vol. II, page 204 of his *Collected Papers*). It has also become known under the names "parity phenomenon" and "parity problem". Some twenty-five years afterward Bombieri [1] went much further along these lines. He showed that, given a "linear" sieve problem with optimal level of distribution, one could say that the contribution to the sequence, in a certain normalized sense, coming from the integers with r prime factors was the same for each odd r and was the same for each even r, and of course, though these two non-negative numbers could be quite different, they total twice the amount that was the expectation for each.

So, in retrospect, one sees now the upper bound and the lower bound are equally unsuccessful. The fact that the lower bound just misses barely crushes the hope of getting prime numbers directly from the sieve. But the fact that we only just miss improving the factor two in the upper bound is equally tantalizing and is related to a much larger story.

There is only time to touch the beginnings of that story, and for this we need to move next into the Analytic Room. The people who live in the Analytic Room also like to count prime numbers but they try to do this using properties of generating functions defined by Euler products and called L-functions. Since Euler is only three hundred years old some of the analytic people like to make fun of the old-fashioned methods used by the elementary people. It turns out that there are crucial facts about counting primes that are hidden in the location of the zeros of these functions. One tries to find regions of the complex plane where there are no such zeros, and an almost universal principle that guides the argument is that these guys are anti-social: the possible existence of two of them in close proximity is not so hard to disprove. The difficult case, which happens for real zeros of real character Dirichlet L-functions, occurs when a badly placed zero has no obvious companion nearby to help rule out its existence. It can be shown without too much effort that a slight improvement of the constant two in the sieve upper bound would disprove the possible existence of any such "exceptional" zeros. We know by the above counterexamples that this is too much to hope for. But what about other arguments? The Riemann Hypothesis would more than suffice. How about something we can actually prove?

There are many ways to try to attack this problem and all of them seem doomed to fail, an amazing number of them by the most narrow of margins. Spilling into the Algebraic Room, we look at the "exceptional" characters whose L-functions could conceivably possess such a bad zero. We nearly find "Selberg counterexamples" in this setting as well. Why does the ring of algebraic integers in the field $Q(\sqrt{-163})$ have unique factorization? Why does the corresponding character (Legendre symbol) look so darn much like the Möbius function? Why does the polynomial $x^2 - x + 41$ take prime values for $0 \leq x \leq 40$? Do these things ever happen again? No? Well, do they almost happen? Why can one just exactly prove, for a family of seemingly irrelevant L-functions, that at least 50 percent of their central values are positive, when almost all might be expected to, and when proving it for 51 percent would suffice to banish the bad guys? Could there be such zeros, such characters?

Sometimes it almost seems as though there is a ghost in the House of Prime Numbers. Perhaps that will be ruled out some day. There are suggestions of a youngster who might do this, one who will come from the Automorphic Room of the house. In the meanwhile, happy-go-lucky prime counters remain temporarily free, see [2], to base some fantastic theorems on either of the two assumptions (that exceptional characters exist or don't exist), whichever one their superstitions dictate.

Further Reading

[1] E. Bombieri, The asymptotic sieve, *Mem. Acad. Naz. dei XL* **1/2** (1975/76), 243–269.

[2] J. Friedlander and H. Iwaniec, The illusory sieve, *Int. J. Number Theory* **1** (2005), 459–494.

[3] H. Iwaniec, Conversations on the exceptional character, *Analytic Number Theory*, Cetraro, Italy, 2002 (A. Perelli and C. Viola, eds.), pp. 97–132, Lecture Notes in Math., 1891, Springer (Berlin), 2006.