Feature

4 October 2017

Riddle of the primes: why do they come in pairs?

Why prime numbers clump the way they do is one of the most maddening problems in mathematics – and thanks to an unlikely hero we're starting to crack it

Ellen Porteus

By Vicky Neale

IT WAS the British mathematician G.H. Hardy who popularised the idea that youthful brains do the best maths. "I do not know of a major mathematical advance initiated by a man past fifty", he wrote in *A Mathematician's Apology*, a lament for the decline of his own creativity that he published in 1940 at the age of 62. "If a man of mature age loses interest in and abandons mathematics, the loss is not likely to be very serious either for mathematics or for himself."

If blooming youth is the rule, Yitang Zhang is a definite exception. For the best part of a decade after completing his PhD, he wasn't even working as a mathematician, instead doing odd accounting jobs around Kentucky. At one point he did a stint working in a Subway fast-food restaurant. When he announced a mathematical breakthrough that had eluded his peers for a couple of centuries, he was 57.

What Zhang made public in 2013 wasn't a proof of the hallowed "twin primes conjecture", but it was a significant step towards one. And even if things haven't quite panned out in the years since, he has inspired work that is promising new insights into the prime numbers, the most beguiling numbers of all.

Primes are those numbers greater than 1 that are divisible only by 1 and themselves. The sequence begins 2, 3, 5, 7, 11, 13, 17, 19 and goes on… well, as long as you like. Primes underpin modern cryptography, keeping your credit card details safe when you shop online. But their true power lies in the crucial role they play in number theory, the branch of mathematics concerned with the properties of whole numbers.

Primes are the fundamental entities from which we make all numbers, because any number that is not prime can be obtained by multiplying other primes together. "It's the same idea as in chemistry, where you might try to understand some complicated compound by understanding the atomic elements which it is made from and how they are joined together," says James Maynard, a mathematician at the University of Oxford.

The fascination with primes goes back at least as far as the ancient Greeks. In *The Elements*, Euclid came up with a beautiful proof that there are infinitely many primes, so there is no largest prime number.

Let's assume for a moment you have a list of all the prime numbers. Multiply all these together, then add 1, and you get a number that, by definition, cannot be divided exactly by any of the primes used to make it: 1 will always be left over as a remainder. Either it is divisible by another prime not on the list, or it is itself prime – so the original list must be incomplete. You can repeat this reasoning with any initial list of primes, so it follows that no finite list of primes contains them all.

"When it comes to theorems about patterns governing prime numbers, things rapidly become distinctly fuzzy"

That's crystal clear – but when it comes to theorems about patterns governing prime numbers, and in particular where they fall on the number line formed by placing all whole numbers in order, things rapidly become distinctly fuzzy.

At the end of the 19th century, Frenchman Jacques Hadamard and Belgian Charles de la Vallée Poussin independently proved what's known as the prime number theorem, which gives an estimate of the number of primes that are smaller than a million, or a trillion, or indeed any value. This theorem tells us that, on average, the primes get more spread out as we go along the number line. That fits neatly with our experience of the primes up to 100, say: the first few, 2, 3 and 5, are squashed up close, whereas there's a big gap between the two biggest primes less than 100: 89 and 97.

But right after 100 we find the primes 101, 103, 107 and 109 all bunched up together. While it is true that bigger primes get more spread out, that's just on average: look closely, and their behaviour is more nuanced.

And that's where the twin primes conjecture comes in. Apart from 2 and 3, there can't be any pairs of consecutive numbers that are both prime – one would have to be an even number, divisible by 2. But as those first primes after 100 suggest, there are many pairs of primes that differ by 2, such as 3 and 5, or 41 and 43, or 107 and 109.

Infinite twins

The twin primes conjecture predicts that, just as there are infinitely many primes, there are infinitely many pairs of these twin primes: our supply will never run out. There are good reasons to think this is the case. The first is that, with the help of computers, we have found many large twin primes. It might be, though, that the computer has found the largest there is. More compellingly, mathematicians have a model to make predictions about how many twin primes there should be up to a given point along the number line. When checked against calculations made by a computer capable of identifying twin primes out into the furthest reaches, where the truly gargantuan numbers live, the model is remarkably accurate – and it predicts there are infinitely many twin primes.

Mathematicians, though, need absolute certainty, a rigorously reasoned argument that leaves no room for doubt, as with Euclid's "proof by contradiction" argument that there are infinitely many primes. Yet even after grappling with the twin primes conjecture for hundreds of years, mathematicians have so far failed to come up with such a proof.

Hence the shock in 2013, when Zhang proved that there are infinitely many pairs of consecutive primes with a gap less than 70 million. By this point Zhang was a lecturer at the University of New Hampshire, but he had published next to nothing, so there was no suggestion that something like this was in the offing. His watertight proof made him a mathematics superstar overnight. He was inundated with job offers from prestigious institutions such as the University of California, Santa Barbara, where he now works.

Even more remarkable was that Zhang's breakthrough exploited an approach that most of the best mathematical minds had ruled out. This "sieve method" started with the ancient Greek mathematician Eratosthenes, who used it as a handy way to shake out prime numbers from the rest. In the case of finding all the primes up to 100, say, it relies on methodically crossing out all the numbers that are not prime. But that is too blunt an instrument to locate particular patterns of primes, so mathematicians have refined their sieving tools in various ways over the centuries.

Sift it out

The sieve of Eratosthenes offers a simple way to find all the primes up to any given number. Cross out 1, which is by definition not a prime. Then cross out all multiples of 2, 3, and progressively multiples of any number not yet crossed out, for example 5 and 7. What you're left with are the primes (circled)

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Just over a decade ago, Daniel Goldston, János Pintz and Cem Yildirim came up with a modified version of the sieve that came tantalisingly close to proving there are infinitely many pairs of primes that differ by at most 16. To make it work, however, they had to assume another unproven conjecture was true. This is a well-established way to make progress, but means the result doesn't amount to a complete proof. Zhang, on the other hand, was able to modify the sieve method so as not to rely on unproven assumptions.

Proving there are infinitely many consecutive primes separated by at most 70 million might sound distinctly unimpressive when the goal is 2, but 70 million is a lot less than infinity. What's more, this was the first time anyone had managed to prove there are infinitely many primes with a gap less than some fixed finite number. "Just to have a number was extraordinary," says Andrew Granville, a number theorist at the University of Montreal and University College London. "Everybody had tried to find a proof along these lines and I really didn't think it was possible."

As soon as the proof was published, mathematicians scrambled to understand Zhang's approach. The limit of 70 million was not the best that his argument would give, so others set about tightening up the details of the proof. The charge was led by Scott Morrison of the Australian National University, and subsequently Fields medallist Terry Tao of the University of California, Los Angeles, who started an online Polymath collaboration to tackle the problem more systematically. The idea with Polymath projects is that all contributors can work on an unsolved problem, collaborating entirely in public on blogs and wikis.

"Zhang's watertight proof made him a mathematics superstar overnight"

It worked beautifully in this case: within months the collaboration was able to prove that there are infinitely many pairs of primes where the gap is less than or equal to 4680. But then progress dried up. The Polymathematicians had squeezed the best they could out of Zhang's argument, and needed new tools to go further.

It took a fresh perspective from Maynard, then a postdoc at the University of Montreal in Canada, to make the gap shrink again. Revisiting the work of Goldston, Pintz and Yildirim, he found a new way to use a sieve that was both simpler than Zhang's and gave a better result: there are infinitely many pairs of primes that differ by at most 600.

By April 2014, the Polymath project was back in the game and, using the new method, brought the gap down from 600 to less than or equal to 246. That is a huge improvement on 70 million, never mind infinity. And that, for now, is the state of the art: all the methods that got us this far have come up against the mathematical equivalent of a brick wall.

The trouble lies in the definition of a prime number, and the way sieve theory works. A prime number always has just one prime factor, namely itself. Sieve theory struggles when it's only looking for numbers with an odd number of prime factors. "It is sort of like a radar that's trying to scan for prime numbers but it gets lots of false positives," says Maynard. "You can't tell which bleeps come from primes and which come from numbers that look like primes but actually have two or four prime factors." This is what mathematicians call the parity problem, and right now there seems to be no way around it.

But Maynard has a sniff of something promising: a recent breakthrough, which gives a way of zooming in from the average behaviour of numbers across long intervals of the number line to work out patterns over shorter intervals. That was long thought to be incredibly hard, if not impossible, but in 2015, Kaisa Matomäki of the University of Turku in Finland and Maksym Radziwiłł, now at McGill University in Montreal, Canada, were able to do exactly that. "They showed that almost all the time, if you just pick some zoomed-in place, you'll get numbers with an even number of prime factors and numbers with an odd number of prime factors," says Maynard. "It's a technical result that is very exciting for us, because these nuts and bolts can often be applied in other areas."

Indeed, Tao has already used the insight to solve Chowla's conjecture, a "baby version" of the twin primes conjecture, which was created as a sort of stepping stone towards that proof. He looked at the sequence of numbers starting with 1 × 3, 2 × 4, 3 × 5, 4 × 6, 5 × 7, and showed that a number in this sequence is equally likely to have an odd number or an even number of prime factors.

Neither of these developments directly deals with the twin primes conjecture and, although Granville was "shocked" to see Matomäki and Radziwiłł's result, he is yet to be convinced it will help with the twin primes conjecture. "It's not at all clear how this will play out," he says.

Such is the nature of maths: you never quite know when painstakingly slow progress behind the scenes will suddenly fall into place for a big breakthrough. For Maynard, however, the signs are at least now hopeful. "The mere fact that people have handled the parity problem in contexts that are not too far away from the twin primes conjecture makes me optimistic". The mysteries of the twins could soon be up for grabs, but it might take another left-field hero like Zhang to make the breakthrough.

Prime problems

The riddle of the never-ending pairs is not the only mystery of the prime numbers

Goldbach's conjecture

This is the prediction that every even number above 4 can be written as a sum of two odd prime numbers – for example, 10 = 3 + 7, and 78 = 31 + 47. Proposed by Christian Goldbach in 1742, it remains unproven.

Infinite Germains

A Germain prime, named after Sophie Germain, is one that gives another prime if you double it and add 1. For example, 29 is prime, and (29 x 2) + 1 = 59 is also prime, so 29 is a Germain prime. Mathematicians expect that there are infinitely many Germain primes, but no one can prove it.

The Riemann hypothesis

In 1859, Bernhard Riemann put forward an idea about where the Riemann zeta function takes the value zero. Proving this conjecture would reveal more about the distribution of the primes. It is one of the Clay Mathematics Institute's seven Millennium Problems – prove Riemann's idea and you win $1 million.