

Polymath projects

Massively collaborative online mathematics

Terence Tao

University of California, Los Angeles

Breakthrough Mathematics Symposium

November 10, 2014

Massively collaborative research projects, such as the discovery of the Higgs Boson ([Fundamental Physics Prize 2012](#))...



ELSEVIER

Physics Letters B

Volume 716, Issue 1, 17 September 2012, Pages 1–29



Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC ☆

ATLAS Collaboration*

G. Aad⁴⁸, T. Abajyan²¹, B. Abbott¹¹¹, J. Abdallah¹², S. Abdel Khalek¹¹⁵, A.A. Abdelalim⁴⁹, O. Abdinov¹¹, R. Aben¹⁰⁵, B. Abi¹¹², M. Abolins⁸⁸, O.S. AbouZeid¹⁵⁸, H. Abramowicz¹⁵³, H. Abreu¹³⁶, B.S. Acharya^{164a, 164b}, L. Adamczyk³⁸, D.L. Adams²⁵, T.N. Addy⁵⁶, J. Adelman¹⁷⁶, S. Adomeit⁹⁸, P. Adragna⁷⁵, T. Adye¹²⁹, S. Aefsky²³, J.A. Aguilar-Saavedra^{124b, a}, M. Agustoni¹⁷, M. Aharrouche⁸¹, S.P. Ahlen²², F. Ahles⁴⁸, A. Ahmad¹⁴⁸, M. Ahsan⁴¹, G. Aielli^{133a, 133b}, T. Akdogan^{19a},

[+ Show more](#)

(There are 3, 171 coauthors in the full list.)

... or the Human Genome Project ([Breakthrough Prize in Life Sciences 2013](#)), are now commonplace in the physical and life sciences.

nature

International weekly journal of science

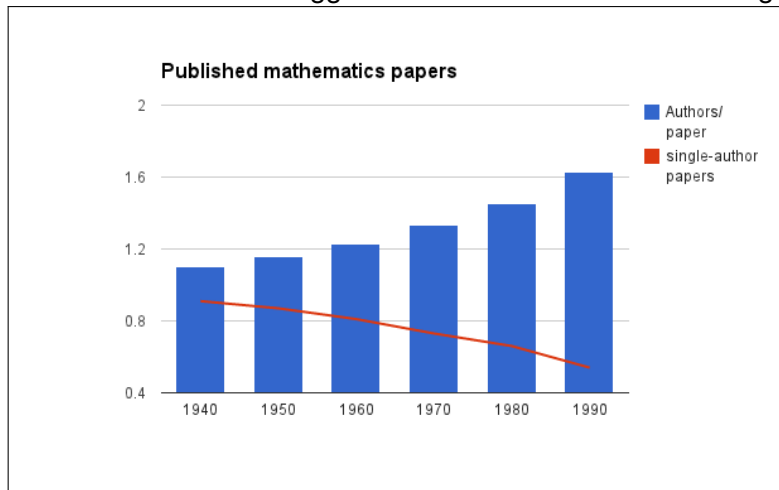
Nature **409**, 860-921 (15 February 2001) | doi:10.1038/35057062; Received 7 December 2000; Accepted 9 January 2001

Initial sequencing and analysis of the human genome

International Human Genome Sequencing Consortium Eric S. Lander¹, Lauren M. Linton¹, Bruce Birren¹, Chad Nusbaum¹, Michael C. Zody¹, Jennifer Baldwin¹, Keri Devon¹, Ken Dewar¹, Michael Doyle¹, William FitzHugh¹, Roel Funke¹, Diane Gage¹, Katrina Harris¹, Andrew Heaford¹, John Howland¹, Lisa Kann¹, Jessica Lehoczy¹, Rosie LeVine¹, Paul McEwan¹, Kevin McKernan¹, James Meldrim¹, Jill P. Mesirov¹, Cher Miranda¹, William Morris¹, Jerome Naylor¹, Christina Raymond¹, Mark Rosetti¹, Ralph Santos¹, Andrew Sheridan¹, Carrie Sougnez¹, Nicole Stange-Thomann¹, Nikola Stojanovic¹, Aravind Subramanian¹ & Dudley Wyman¹ for Whitehead Institute for Biomedical Research, Center for Genome Research; Jane Rogers², John Sulston², Rachael

(There are 237 coauthors in the full list, some representing institutes.)

But mathematics has lagged behind the sciences in this regard.



(Source: J. Grossman, Patterns of Collaborations in Mathematical Research, SIAM News Vol. 35 No. 9, Nov. 2002)

Even today, many mathematical breakthroughs come from individual mathematicians working in isolation for long periods of time. For instance:

- **Andrew Wiles** secretly worked for seven years before revealing his proof of Fermat's last theorem in 1993 (but then collaborated with **Richard Taylor** ([Mathematics Breakthrough Prize 2014](#)) to repair a gap in the first version of this proof).
- **Grisha Perelman** also worked alone for years before releasing his preprints proving the Poincaré conjecture and geometrisation conjectures in 2002-2003.
- **Yitang Zhang** worked for three years by himself before proving in 2013 that bounded gaps between primes occur infinitely often.

But many major mathematical achievements have also been made collaboratively. For instance:

- For much of the twentieth century, **Nicolas Bourbaki** (a pseudonym for the collaborative efforts of dozens of eminent mathematicians) systematically developed the rigorous foundations of modern mathematics in an influential series of textbooks.
- The **classification of finite simple groups** (also known as “the Enormous Theorem”) spans tens of thousands of pages in several hundred journal articles by over a hundred authors, from approximately 1955 to 2004. (A more focused collaboration by Gorenstein, Lyons, Solomon, Aschbacher, and Smith have nearly completed a “second generation” proof of this theorem.)

- In the sciences, there are many tasks that we know how to do **in theory**, but not necessarily **in practice**. In such cases, the injection of massive resources (physical, financial, or human) can make a decisive difference.
- But often in mathematics, the problem is that we don't know how to do a task, *even in theory*. For instance, the **Riemann hypothesis** - one of the most important unsolved problems in mathematics - would likely remain out of reach in the near future even if we threw huge amounts of money, computer power, and people at the problem, because we don't have a plausible theoretical approach to the problem yet.

Nevertheless, in January 2009, **Timothy Gowers** asked whether mathematical research breakthroughs could be “crowdsourced” through massive online collaboration.

Gowers's Weblog

Is massively collaborative mathematics possible?

Of course, one might say, there are certain kinds of problems that lend themselves to huge collaborations. One has only to think of the proof of the classification of finite simple groups, or of a rather different kind of example such as a search for a new largest prime carried out during the downtime of thousands of PCs around the world. But my question is a different one. What about the solving of a problem that does not naturally split up into a vast number of subtasks? Are such problems best tackled by n people for some n that belongs to the set $\{1, 2, 3\}$? (Examples of famous papers with four authors do not count as an interesting answer to this question.)

It seems to me that, at least in theory, a different model could work: different, that is, from the usual model of people working in isolation or collaborating with one or two others. Suppose one had a forum (in

Instead of having a small number of mathematicians work in isolation for extended periods of time before releasing carefully polished reports, the paradigm Gowers proposed was instead to have a very large number of small contributions coming from a diverse group of mathematicians, each working for a short period of time before sharing their (not necessarily complete or correct) findings.

(iii) Different people have different characteristics when it comes to research. Some like to throw out ideas, others to criticize them, others to work out details, others to re-explain ideas in a different language, others to formulate different but related problems, others to step back from a big muddle of ideas and fashion some more coherent picture out of them, and so on. A hugely collaborative project would make it possible for people to specialize. For example, if you are interested in the problem and like having slightly wild ideas but are less keen on the detailed work of testing those ideas, then you can just suggest the ideas and hope that others will find them interesting enough to test or otherwise respond to.

In short, if a large group of mathematicians could connect their brains efficiently, they could perhaps solve problems very efficiently as well.

After a few weeks of online discussion, the first **Polymath project** was launched, using off-the-shelf online tools such as blogs and wikis.

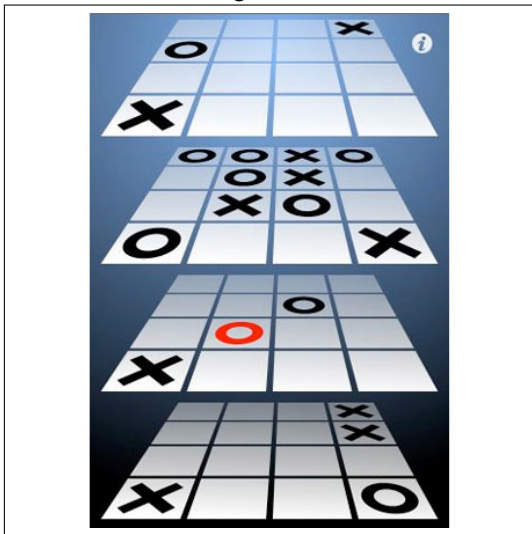
Gowers's Weblog

A combinatorial approach to density Hales-Jewett

Here then is the project that I hope it might be possible to carry out by means of a large collaboration in which no single person has to work all that hard (except perhaps when it comes to writing up). Let me begin by repeating a number of qualifications, just so that it is clear what the aim is.

1. It is *not* the case that the aim of the project is to find a combinatorial proof of the density Hales-Jewett theorem when $k = 3$. I would love it if that was the result, but the actual aim is more modest: it is *either* to prove that a certain approach to that theorem (which I shall soon explain) works, *or* to give a very convincing argument that that approach cannot work. (I shall have a few remarks

Polymath1 was interested in finding a new, and purely combinatorial proof, of the **density Hales-Jewett theorem**. This combinatorial theorem can be explained in terms of multi-dimensional tic-tac-toe games.



- Roughly speaking, the theorem says that such games cannot end in a draw if the dimension of the game is large enough. (In fact, the game will necessarily stop with a winner before even a tiny fraction of the positions are filled.)
- This may seem like a somewhat frivolous mathematical statement, but it turns out to be quite a powerful combinatorial tool, being one of the deepest results of **Ramsey theory**. For instance, it can be used to prove a result of Ben Green and myself that the primes contain arbitrarily long arithmetic progressions.
- Until the Polymath1 project, the only known proof of this theorem was a difficult argument of Furstenberg and Katznelson using advanced methods from ergodic theory.

After over a month of intense activity by over two dozen mathematicians (ranging from Fields Medalists to amateur enthusiasts), and nearly a thousand comments across more than a dozen blog posts, Polymath1 found a simple combinatorial proof of the density Hales-Jewett theorem.

Annals of Mathematics **175** (2012), 1283–1327
<http://dx.doi.org/10.4007/annals.2012.175.3.6>

A new proof of the density Hales-Jewett theorem

By D. H. J. POLYMATH

The Hales-Jewett theorem asserts that for every r and every k there exists n such that every r -colouring of the n -dimensional grid $\{1, \dots, k\}^n$ contains a monochromatic combinatorial line. This result is a generalization of van der Waerden's theorem, and it is one of the fundamental results of Ramsey theory. The theorem of van der Waerden has a famous density version, conjectured by Erdős and Turán in 1936, proved by Szemerédi in 1975, and given a different proof by Furstenberg in 1977. The Hales-Jewett theorem has a density version as well, proved by Furstenberg and Katznelson in 1991 by means of a significant extension of the ergodic techniques that had been pioneered by Furstenberg in his proof of Szemerédi's theorem. In this paper, we give the first elementary proof of the theorem of

Since then, there have been eight further “official” Polymath projects, as well as some further polymath-inspired projects.

Existing polymath projects

- **Polymath1**: New proofs and bounds for the density Hales-Jewett theorem. Initiated Feb 1, 2009; research results have now been published.
- **Polymath2**: Must an “explicitly defined” Banach space contain c_0 or l_p ? Initiated Feb 17, 2009; attempts to relaunch via wiki, June 9 2010.
- **Mini-polymath1**: Solving Problem 6 of the 2009 International Mathematical Olympiad. Initiated July 20, 2009; five proofs obtained so far.
- **Polymath3**: The polynomial Hirsch conjecture. Proposed July 17, 2009; launched, September 30, 2010.
- **Polymath4**: A deterministic way to find primes. Proposed July 27, 2009; launched Aug 9, 2009. Research results have been accepted for publication.
- **Polymath5**: The Erdős discrepancy problem. Proposed Jan 10, 2010; launched Jan 19, 2010.
- **Mini-polymath2**: Solving Problem 5 the 2010 International Mathematical Olympiad. Proposed Jun 12, 2010; launched and solved, Jul 8 2010.
- **Polymath6**: Improving the bounds for Roth’s theorem. Proposed Feb 5, 2011.
- **Mini-polymath3**: Solving a problem from the 2011 International Mathematical Olympiad. Proposed Jun 9, 2011; launched and solved, Jul 19, 2011.
- **Mini-polymath4**: Solving a problem from the 2012 International Mathematical Olympiad. Proposed, Jun 3, 2012; launched, July 12 2012.
- **Polymath7**: Establishing the Hot Spots conjecture for acute-angled triangles. Proposed, May 31st, 2012; launched, Jun 8, 2012.
- **Polymath8**: Improving the bounds for small gaps between primes. Proposed, June 4, 2013; launched, June 4, 2013.
- **Polymath9**: exploring Borel determinacy-based methods for giving complexity bounds. Proposed, Oct 24, 2013; launched, Nov 3, 2013.

Polymath-like projects

- “Theory of singularities” research attempt [🔗](#) in the form of a TiddleSpace wiki
- Attempt to prove that certain categories are cartesian closed [🔗](#) in the form of a TiddleSpace wiki
- Mathematics research projects [🔗](#)
- Scott Aaronson’s “philomath project”: “Sensitivity vs. Block sensitivity [🔗](#)” (see also this Math Overflow question [🔗](#)). Launched Jul 13, 2010.

Some projects fizzled out, but others have been quite successful. The most recent successfully concluded project was the **Polymath8 project** on bounded gaps between primes, building upon the breakthrough in 2013 of Yitang Zhang.

Annals of Mathematics **179** (2014), 1121–1174
<http://dx.doi.org/10.4007/annals.2014.179.3.7>

Bounded gaps between primes

By YITANG ZHANG

It is proved that

$$\liminf_{n \rightarrow \infty} (p_{n+1} - p_n) < 7 \times 10^7,$$

where p_n is the n -th prime.

Our method is a refinement of the recent work of Goldston, Pintz and Yıldırım on the small gaps between consecutive primes. A major ingredient of the proof is a stronger version of the Bombieri-Vinogradov theorem that is applicable when the moduli are free from large prime divisors only, but it is adequate for our purpose.

- Yitang Zhang showed there were infinitely many pairs of primes whose difference were bounded by 70,000,000.
- If this bound could be lowered to 2, this would prove the notorious **twin prime conjecture**.
- He knew that his bound of 70,000,000 could be improved, but left this task to others.

$$(1.5) \quad \liminf_{n \rightarrow \infty} (p_{n+1} - p_n) < 7 \times 10^7.$$

The bound (1.5) results from the fact that the set \mathcal{H} is admissible if it is composed of k_0 distinct primes, each of which is greater than k_0 , and the inequality

$$\pi(7 \times 10^7) - \pi(3.5 \times 10^6) > 3.5 \times 10^6.$$

This result is, of course, not optimal. The condition $k_0 \geq 3.5 \times 10^6$ is also crude and there are certain ways to relax it. To replace the right side of (1.5) by a value as small as possible is an open problem that will not be discussed in this paper.

Almost immediately, mathematicians began to spontaneously improve this bound online.

According to WolframAlpha (tinyurl.com/kb363qr) Zhang's argument as is actually gives a gap size of 63,374,611. – [Mark Lewko](#) May 21 '13 at 20:37

A poor man's improvement on Zhang's result:
there are infinitely many prime gaps less than 60
million

T. S. Trudgian*

May 29, 2013

I just can't resist: there are infinitely many pairs
of primes at most 59470640 apart [May 30, 2013](#)

Posted by Scott Morrison in [Number theory](#), [polymath](#).

Tags: [admissible-set](#), [polymath8](#), [primes](#), [zhang](#)



Terence Tao May 30, 2013 +8 +1

Edit ✕

I just posted a comment there in which I get an improvement to 58,885,998 conditional on a plausible claim, and with a bit of numerical calculation one should be able to obtain unconditional improvements on this bound, but I'll leave it to someone else to jump in.

11. Scott Morrison – May 30, 2013

My computer just finished looking through values of m that might work in Richard's sequence described in Terry's comment #8 above. As it turns out $m=36716$ is the least m so that that sequence is admissible. The biggest difference in that sequence is $2 p_{\lfloor 36716+k_0/2-1 \rfloor} = 57,554,086$, giving a new best upper bound on prime gaps.

17. Scott Morrison – May 31, 2013

I made a little mathematica notebook with the relevant formulas, and found the best value of k_0 that works with Zhang's method.

The least k_0 for which there is some l_0 so $\omega(k_0, l_0, 1/1168) > 0$ is 2947442, (which works for $l_0 = 151$ and no other). For reference ω at these parameters is $3.998282544333942 * 10^{-17790133}$; pretty tiny!

I haven't re-run the search for the best m with this new value of k_0 , but right away we get a new bound on prime gaps of $2p_{36716+2947442/2-1} = 48112378$, so hooray, we're below 50 million. :-)

24. Scott Morrison – May 31, 2013

Oh, and $m=30798$ just came in for $k_0 \leq 2,947,442$, giving the infinitesimal improvement to $2p_{30798+\lceil 2618607/2 \rceil + 1} = 42,342,946$.

30. Scott Morrison – June 2, 2013

Happily finding the least m that works in Richard's sequence is now reasonably fast. For $k_0 = 866, 805$, $m=11651$ is the first that works, giving the latest bound of

$$p_{11651+\lfloor 866,805/2 \rfloor - 1} + p_{11651+\lfloor 866,806/2 \rfloor - 1} = 13,008,612.$$

50. Scott Morrison – June 3, 2013

And indeed, with $l_0 = 291.7$, Terry's #47 allows the optimal $k_0 = 341,640$.

I haven't updated m yet, but this already gives us a prime gap of 4,982,086.

At this point, it became clear that these efforts should be run more systematically, and the **Polymath8** project was launched.

June 4, 2013

Polymath proposal: bounded gaps between primes

Filed under: [planning](#), [polymath proposals](#) — Terence Tao @ 4:31 am [Edit This](#)

Two weeks ago, Yitang Zhang announced [his result](#) establishing that bounded gaps between primes occur infinitely often, with the explicit upper bound of 70,000,000 given for this gap. Since then there has been a [flurry of activity](#) in reducing this bound, with the current record being 4,802,222 (but likely to improve at least by a little bit in the near future).

It seems that this naturally suggests a Polymath project with two interrelated goals:

1. Further improving the numerical upper bound on gaps between primes; and
2. Understanding and clarifying Zhang's argument (and other related literature, e.g. the work of Bombieri, Fouvry, Friedlander, and Iwaniec on variants of the Elliott-Halberstam conjecture).

Polymath8 naturally split into smaller groups, each working to optimise one of the aspects of Zhang's argument. The bound on gaps between primes improved at a rapid rate, often multiple times in a day.

Timeline of prime gap bounds

Date	ϖ or (ϖ, δ)	k_0	H	Comments
Aug 10 2005		6 [EH]	16 [EH] ([Goldston-Pintz-Yildirim])	First bounded prime gap result (conditional on Elliott-Halberstam)
May 14 2013	1/1,168 (Zhang)	3,500,000 (Zhang)	70,000,000 (Zhang)	All subsequent work (until the work of Maynard) is based on Zhang's breakthrough paper.
Jun 23		1,466 (Paldi/Harcos)	12,006 (Engelsma)	An improved monotonicity formula for $G_{k_0-1, \delta}$ reduces κ_3 somewhat
Jun 24	$(134 + \frac{2}{3})\varpi + 28\delta \leq 1?$ (v08ltu) $140\varpi + 32\delta < 1?$ (Tao) 1/88?? (Tao) 1/74?? (Tao)	1,268? (v08ltu)	10,206? (Engelsma)	A theoretical gain from rebalancing the exponents in the Type I exponential sum estimates
Jun 25	$116\varpi + 30\delta < 1?$ (Fouvry-Kowalski-Michel-Nelson/Tao)	1,346? (Hannes) 602?? (Trevino) 1,007? (Hannes)	10,876?? (Engelsma) 3,642 (Engelsma) 7,860?? (Engelsma)	Optimistic projections arise from combining the Graham-Ringrose numerology with the announced Fouvry-Kowalski-Michel-Nelson results on d_3 distribution

After almost a year of effort, with thousands of comments on 37 different blog posts by dozens of mathematicians, covering many relevant areas of mathematical expertise, we found a number of new techniques and insights in the subject, and also lowered the bound on gaps between primes to 246.

Polymath Research in the Mathematical Sciences 2014, 1:12
<http://www.resmathsci.com/content/1/1/12>

 Research in
the Mathematical Sciences
a SpringerOpen Journal

Variants of the Selberg sieve, and bounded intervals containing many primes

DHJ Polymath

Correspondence:
tao@math.ucla.edu

<http://michaelnielsen.org/polymath1/index.php>

Abstract

For any $m \geq 1$, let H_m denote the quantity $\liminf_{n \rightarrow \infty} (p_{n+m} - p_n)$. A celebrated recent result of Zhang showed the finiteness of H_1 , with the explicit bound $H_1 \leq 70,000,000$. This was then improved by us (the Polymath8 project) to $H_1 \leq 4680$, and then by Maynard to $H_1 \leq 600$, who also established for the first time a finiteness result for H_m for $m \geq 2$, and specifically that $H_m \ll m^3 e^{4m}$. If one also assumes the Elliott-Halberstam conjecture, Maynard obtained the bound $H_1 \leq 12$, improving upon the previous bound $H_1 \leq 16$ of Goldston, Pintz, and Yıldırım, as well as the bound $H_m \ll m^3 e^{2m}$. In this paper, we extend the methods of Maynard by generalizing the Selberg sieve further and by performing more extensive numerical calculations. As a consequence, we can obtain the bound $H_1 \leq 246$ unconditionally and $H_1 \leq 6$ under the assumption

Not all Polymath projects have been as successful. Factors that worked in our favor for Polymath8 included

- intense interest in the problem from the wider mathematical community;
- a modular structure to the problem, so that people could contribute to one aspect without necessarily being expert with all other aspects;
- a easily comprehended way to measure progress (and one which was guaranteed to terminate!);
- the generous amounts of time and effort put in by the Polymath participants and moderators, and their willingness to openly discuss mathematics (and occasionally to make mistakes) in public.

Up until now, Polymath projects have used “off-the-shelf” free online tools to communicate: they are not ideal in some respects, but they are easy to use, encourage casual participation from people who might not otherwise get involved, and require little technical support. Fancier, more customised software might be nice to have, but at this stage does not seem to be crucial.



Polymath projects tend to be very good, and very fast, at getting around minor technical “speed bumps” such as requiring an obscure research article or substantial computer processing power, tracking down a little-known piece of mathematical folklore, performing a tricky computation, or coding up a specific software task.

Narrow admissible tuples

A k -tuple is a set of k integers. A k -tuple S is *admissible* if, for every prime p , there is a residue class (mod p) that does not intersect the largest and smallest elements) of an admissible k -tuple.

This is an alpha-release with limited functionality. Please send bug reports and feature requests to Andrew Sutherland at

For a complete list of current upper bounds as an ascii text file, [click here](#).

Showing first 100 records with $k \geq$

k	$\geq H(k)$	data	submitter	comment	date
4000	36610	k-tuple	Sutherland		2013-06-27
4001	36622	k-tuple	Sutherland		2013-06-27
4002	36644	k-tuple	Sutherland		2013-06-28
4003	36652	k-tuple	xfxie		2013-06-29
4004	36660	k-tuple	xfxie		2013-06-29
4005	36666	k-tuple	Sutherland		2013-07-02
4006	36672	k-tuple	Sutherland		2013-07-02
4007	36682	k-tuple	xfxie		2013-06-29
4008	36694	k-tuple	xfxie		2013-06-29

They also excel at locating people with relevant expertise who could help with the project, but would never even have heard about the problem, or been considered as someone to contact, if the project were done more traditionally.

4 March, 2014 at 8:38 am

Ignace Bogaert

(edit)

Hello! I have only been (quietly) following this blog for the last two months. My normal research has got nothing to do with number theory, but rather with solving large



linear systems, usually by means of Krylov subspace methods. Since this and previous posts have made the link between quadratic forms and bounded gaps between primes, I have gotten interested in applying Krylov subspaces for the problem of computing M_k . (I know that other regions than the unit simplex allow better values for k to be found, but I wanted to start with a relatively doable problem.)

I have tried a different (and hopefully interesting) way to compute M_k , based on associating a linear operator with the relevant quadratic form. A description of the

However, there are still significant limitations to the Polymath paradigm:

- Polymath projects only seem to work well when there is a large number of people who have enough mathematical background to understand at least part of the problem, and comment on it. Problems that require a lot of very specialised and technical mathematical expertise to even comprehend are poor candidates for a Polymath approach. (However, Polymath-style “online reading seminars” of difficult research papers have been somewhat successful.)
- We do not have a good system for formally recognising one’s contribution to a Polymath project; research papers from such a project are usually written under a pseudonym (we’ve used “D.H.J. Polymath”, inspired by the first Polymath project on the Density Hales-Jewett problem). This is a particular issue for early career mathematicians considering devoting substantial research time to a Polymath project.

- Successful Polymath projects have invariably needed a project leader to moderate and guide the discussion, and to generally keep the momentum going. This can be *very* time consuming, and only a few people have stepped forward to do so thus far (and so we only average about one Polymath project a year).
- Polymath projects have only made progress on problems where there was already some number of promising ways to make progress, for instance by trying to adapt some arguments already in the literature. For the truly difficult mathematical problems, where all known methods have failed and some genuinely new idea or insight is needed, it doesn't look as if a Polymath project would get much further than an individual mathematician would (although one may perhaps get a better insight as to *why* all known methods fail).

- Still, we think there is a significant niche role for Polymath-style projects to play in the future of mathematical research, and perhaps one day we will see truly large-scale mathematics collaborations to rival those in the physical and life sciences.
- Thanks for listening!