

# Research Statement

I am interested in combinatorial questions with a number-theoretic flavour, and also in number-theoretic questions with a combinatorial flavour. I believe that strong numerical evidence can sometimes be translated into explicit constructions if one tries the right things and has a fair amount of luck. This conviction stems from personal experiences in my research career.

## Dissertation and Postgraduate Research

A substantial part of my thesis is related to problems in discrepancy theory, and I had the privilege of working under the supervision of Professor József Beck, one of the subject's pioneers. Discrepancy theory investigates irregularities in the distribution of points in some ambient space. Quite often, coming up with a perfectly equitable distribution turns out to be as difficult as the age-old problem of pleasing everyone. A classic theorem of Schmidt [18] states that no matter how  $N$  points are placed inside the unit square, there will always be a rectangle with sides parallel to the co-ordinate axes and containing at least  $(\log N)/700$  points more or fewer than what one would expect from the area of the rectangle. Since this result happens to be the best possible (except for the constant), we say that the discrepancy of axis-parallel rectangles is  $\Theta(\log N)$ . In dimensions three and higher, the exponent of  $\log N$  in the best known upper and lower bounds are off by a factor of 2, and it is an outstanding open problem to close this gap.

Discrepancy theory also has a combinatorial facet. Suppose we are required to colour each integer in the set  $\{0, 1, \dots, N\}$  red or blue, in such a way that all arithmetic progressions have small discrepancy (i.e., small difference in the number of reds and blues). How well can this be done? In 1964, Roth [17] showed that there always exists an arithmetic progression with discrepancy of the order of  $N^{1/4}$ . In 1996, Matoušek and Spencer [14], building upon the work of Sárközy and Beck, showed that Roth's estimate was sharp up to constants. This is a pleasant contrast with the state-of-the-art in Ramsey theory, where the upper and lower bounds on the length of monochromatic arithmetic progressions are nowhere close to each other.

But all is not rosy. A related and notoriously difficult question is the *unbounded discrepancy problem*, raised by Erdős in the 1930s and given a front seat on his famous list of open problems in combinatorics nearly five decades later (see [9]): Is the discrepancy of homogeneous arithmetic progressions (the ones containing 0) bounded or unbounded? Essentially nothing is known to this day.

As a tractable alternative, one could ask the same question about quasi-arithmetic progressions, which are obtained by taking the integer parts of consecutive multiples of real numbers. For example, the multiples of  $\sqrt{13} = 3.605\dots$  yield the progression  $\{0, 3, 7, 10, 14, \dots\}$ . It turns out to be impossible to balance even a significant fraction of these quasi-arithmetic progressions. In 1986, Beck [3] showed that for almost every real number, the discrepancy of the associated progression tends to infinity with  $N$  at least as

fast as the function  $\log^* N$ , the inverse of the tower function. Using tools from harmonic analysis and inspired by techniques in a paper of Roth (see [16]), I have been able to improve on this result, and show (see [23]) that the discrepancy grows at least as fast as a power of  $\log N$  for almost every real number. I can also show that given any 2-colouring, there exists a quasi-arithmetic progression with discrepancy at least  $N^{1/6}/50$ . These results remain valid even for partial colourings of positive density, unlike homogeneous arithmetic progressions.

Quasi-arithmetic progressions are obtained by subtracting the fractional parts from the sequence of multiples of a given real number. The fractional parts themselves have interesting properties. The well-known three distance theorem, proved independently by several authors (see [20] and [21]) in the 1950s, states that there are at most three distinct gaps between consecutive elements of the first  $n$  multiples of any fixed real number. Higher dimensional generalisations of this theorem are known (see [6] and [7]), but none have the same punchline. Since it is possible to develop the theory of continued fractions starting from the three distance theorem, analogous results in higher dimensions may shed some light on simultaneous diophantine approximability, a fundamental problem in number theory (see [8]).

I have obtained (see [24]) a generalisation of the three distance theorem that is faithful to the original. The theorem can be interpreted as a statement about champions in a tournament. The players are edges between pairs of fractional parts of multiples of the given real number, the matches are between overlapping edges, and when two edges play, the shorter edge wins. According to the three distance theorem, the set of lengths of champions (undefeated edges) has size at most three. In higher dimensions, we consider a vector of real numbers. Two edges play if their projections overlap along any of the coordinate axes, and the shorter edge wins. Under the Euclidean metric, the set of lengths of champions has size at most 11 in the plane, and at most 290 in three dimensions.

A fundamental result in Ramsey theory is the existence of the Hales-Jewett number  $HJ(n)$ . This is the least dimension  $d$  such that no matter how the cells of the  $n^d$  hypercube are 2-coloured, there exists a monochromatic line. Consequently, for each  $n$ , there exists  $d$  such that no game of  $d$ -dimensional  $n$ -in-a-line tic-tac-toe can end in a draw. It follows that the first player has a winning strategy, since if the second player had one, the first player could steal it, and the extra move can only help. While the game-theoretic threshold could be much lower, nothing more is known.

József Beck, Wesley Pegden and I (see [4]) recently obtained an exponential lower bound on  $HJ(n)$ , improving the linear lower bound in the original paper of Hales and Jewett (see [12]). However, there still remains an astronomical gap between the upper and lower bounds, and it would be nice to have a “reasonable” upper bound. The best upper bound on  $HJ(4)$ , due to Shelah [19], is an exponential tower of 2’s of height 24. On the bright side, we do know that  $HJ(3) = 3$ .

A game-theoretic variant of the unbounded discrepancy problem would be to consider two players, Maker and Breaker, who take turns colouring the integers from 0 to  $N$  with their own colours. Maker wins if his/her lead on some homogeneous arithmetic progression exceeds a pre-specified target, and Breaker wins otherwise. Assuming both players play perfectly, who wins? Clearly, the answer depends on the target. I can prove that given any  $\varepsilon > 0$ , Maker wins if the target is below  $N^{1/2-\varepsilon}$  and Breaker wins if the target is above  $N^{1/2+\varepsilon}$  for sufficiently large  $N$  (depending on  $\varepsilon$ ). The proof makes use of potential functions, a standard tool in combinatorial game theory.

Towards the end of my first year of study at Rutgers, I became aware of an intriguing conjecture. Professor Doron Zeilberger had obtained, a few years ago, compelling numerical evidence that the popular Hanukkah game Dreidel, with two players and  $n$  tokens each, lasts  $O(n^2)$  spins on average. In one of our department seminars, he offered \$25 for settling the conjecture, and an additional \$25 for solving the case of more than two players. Thomas Robinson and I showed (see [15]), with the help of Wald's equation, Markov chains, and ad hoc combinatorial hacks, that Dreidel indeed lasted  $O(n^2)$  spins on average, irrespective of the number of players.

It is well-known, and easy to show, that the largest subset of  $\{1, 2, \dots, n\}$  such that no number divides another has  $\lceil n/2 \rceil$  elements. Sorting the numbers in decreasing order and choosing as many as possible turns out to be the best option. However, this does not generalise. In 1977, Lebensold showed that the largest subset of  $\{1, 2, \dots, n\}$  such that no number divides two others has substantially more than  $2n/3$  elements, in fact at least  $0.672n$  (but fewer than  $0.674n$ ) elements for sufficiently large  $n$ . I was able to show (see [22]) that if no number is allowed to divide  $k$  others, there is a way of exceeding the greedy ratio of  $k/(k+1)$  by at least  $(1/8)k^{-4}$ . I also obtained an asymptotic upper bound of  $1 - (2e^\gamma k \ln k)^{-1}$ , where  $\gamma = 0.5772\dots$  denotes the Euler-Mascheroni constant. The logarithmic factor in the upper bound was recently cancelled by Hegarty [13].

## Future Goals

I believe that researchers must concentrate on their strengths, rather than work on areas that happen to be fashionable or popular at any given time. It is important that the questions under investigation be fundamental: consequences and applications, whether or not they matter, cannot then be too far behind. I am confident that advances in extremal combinatorics, Ramsey theory and discrepancy theory will enhance our understanding of order and structure in the discrete universe.

There are several outstanding open problems in combinatorial discrepancy theory that have delighted us long enough. The unbounded discrepancy problem, Tusnady's problem, the Beck-Fiala conjecture, and the three permutations conjecture are all exciting questions that have either resisted all attack or have not yet been resolved satisfactorily. The three

permutations conjecture is particularly enticing. Given three permutations of  $\{1, 2, \dots, n\}$ , we wish to colour the first  $n$  integers red or blue so that the colouring induced on the permutations has the following property: the number of reds and blues in any interval differ by an absolute constant. For all we know, there is such an absolute constant and its value is as small as 3. The problem is almost trivial for two permutations, since a straightforward matching algorithm yields an upper bound of 1. For three permutations, the best known upper bound is  $O(\log n)$ , due to Bohus [5].

The gap between upper and lower bounds for diagonal Ramsey numbers is still cause for humility in the combinatorics community. Explicit constructions are even more disappointing. The random-colouring argument yields an exponential lower bound on the diagonal Ramsey number, but the best known constructive lower bounds (see [10]) are of the order of  $n^{\log n / \log \log n}$ . Barak, Rao, Shaltiel and Wigderson [2] have recently obtained an algorithmic improvement, but the gap remains substantial.

I have investigated several 2-colourings based on number-theoretic functions. A very promising colouring  $f(i, j)$  is given by the parity of the length of the continued fraction expansion of  $j/i$ . Let  $R^*(n)$  denote the least order of the complete graph that has a clique or independent set of order  $n$  under this colouring. Values of  $R^*(n)$  for small  $n$  appear to exhibit an exponential trend:

$n$	8	9	10	11	12	13	14
$R^*(n)$	29	45	72	89	135	216	360

Since any 2-colouring on  $n$  vertices has at least  $n^2/4 + O(n)$  edges of one colour, one needs a way of characterising edge sets that correspond to cliques in any proposed pseudorandom construction. The Copeland-Erdős theorem gives many examples of “interval-normal” numbers; we need explicit examples of binary sequences (i.e., 2-colourings of the integers) that *can be proved* “clique-normal”. Perhaps we can reduce the problem to one of exponential sums by formulating the analogue of a Weyl criterion. Such a sequence ought to provide valuable insights on the derandomisation of graph algorithms.

I am also keen on exploring connections between combinatorics, number theory and computer science. There are many deep questions in the theory of algorithms, optimisation and cryptography that I find fascinating. To mention one example, the well-known open problem of efficiently approximating the shortest vector in a lattice up to a polynomial factor gives a fresh perspective on a venerable question studied by Gauss, Dirichlet et al. I will have a lot to learn before I could hope to contribute anything meaningful to this problem, but it is on my shortlist of future projects.

## References

- [1] N. Alon and J. H. Spencer, *The Probabilistic Method*. Academic Press, New York, 1992.

- [2] B. Barak, A. Rao, S. Shaltiel and A. Wigderson, *2-source dispersers for sub-polynomial entropy and Ramsey graphs beating the Frankl-Wilson construction*. Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC '06), 2006.
- [3] J. Beck, *On irregularities of  $\pm 1$ -sequences*. Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber II 195, 1986.
- [4] J. Beck, W. Pegden and S. Vijay, *The Hales-Jewett number is exponential: game-theoretic consequences*. Submitted.  
<http://www.math.rutgers.edu/~sujith/HalesJ.pdf>
- [5] G. Bohus, *On the discrepancy of 3 permutations*. Random Structures and Algorithms 1, 1990.
- [6] N. Chevallier, *Geometrie des suites de Kronecker*. Manuscripta Math. 94, 1997.
- [7] F. R. K. Chung and R. L. Graham, *On the set of distances determined by the union of arithmetic progressions*. Ars Combinatoria 1, 1976.
- [8] H. Davenport and K. Mahler, *Simultaneous Diophantine approximation*. Duke Mathematics Journal 13, 1946.
- [9] P. Erdős, *On the combinatorial problems which I would most like to see solved*. Combinatorica 1, 1981.
- [10] P. Frankl and R. M. Wilson, *Intersection theorems with geometric consequences*. Combinatorica 1, 1981.
- [11] R. L. Graham, B. L. Rothschild and J. H. Spencer, *Ramsey Theory*. Wiley-Interscience Series in Discrete Math., New York, 1980.
- [12] A. W. Hales and R. I. Jewett, *On regularity and positional games*, Trans. Amer. Math. Soc. 106, 1963.
- [13] P. Hegarty, *An improved upper bound on the maximum size of  $k$ -primitive sets*. Integers 6, 2006.
- [14] J. Matoušek and J. H. Spencer, *Discrepancy in arithmetic progressions*. Journal of the American Mathematical Society 9, 1996.
- [15] T. Robinson and S. Vijay, *Dreidel lasts  $O(n^2)$  spins*. Advances in Applied Mathematics 36, 2006.
- [16] K. F. Roth, *On irregularities of distribution*. Mathematika 1, 1954.
- [17] K. F. Roth, *Remark concerning integer sequences*. Acta Arithmetica 9, 1964.
- [18] W. M. Schmidt, *Irregularities of Distribution. VII*. Acta Arithmetica 21, 1972.

- [19] S. Shelah, *Primitive recursive bounds for Van der Waerden numbers*. Journal of the American Math. Soc. 1, 1988.
- [20] V. T. Sós, *On the distribution mod 1 of the sequence  $\{n\alpha\}$* . Ann. Univ. Sci. Budapest, Eotvos Sect. Math. 1, 1958.
- [21] S. Świerczkowski, *On successive settings of an arc on the circumference of a circle*. Fund. Math. 46, 1958.
- [22] S. Vijay, *On the largest  $k$ -primitive subset of  $[1, n]$* . Integers 6, 2006.
- [23] S. Vijay, *On the discrepancy of quasi-progressions*. arXiv:math.CO/0604511, Preprint.
- [24] S. Vijay, *Eleven Euclidean distances are enough*. arXiv:math.CO/0609536, Preprint.