# RIGOROUS EXPERIMENTAL COMBINATORICS: PROJECT DESCRIPTION

*Doron Zeilberger*

Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019. `zeilberg@math.rutgers.edu`
`http://www.math.rutgers.edu/~zeilberg/` .

**Summary of Results from Previous NSF Support: DMS 0401124**

**1.** The current NSF award number is DMS-0401124 for the period 2004-2009 (five summers), totaling $239,023 (including the supplements for graduate student support).

**2.** Its title is: "*Automating Combinatorics*".

**3. Summary of the results of the completed work**

(The numbered references apply to the list of papers written with the NSF support of the above grant, given at section 4. The lettered references are to papers listed in "References Cited".)

It is often the case, in scientific research (and elsewhere), that *the whole is larger than the sum of its parts.* I believe that in the case of my own research, and of my students, the whole is *much* larger than the sum of its parts, not that the parts have anything to be ashamed of.

**The Whole**

My research students and I continued to practice a new research methodology, that can be loosely called *rigorous experimental mathematics.* It has something in common with both "mainstream" experimental mathematics, as preached by the Borwein brothers, David Bailey, Victor Moll, and their collaborators, (see e.g. the masterpiece [BB], and the recent collection [BBCGLM]), and *automated theorem proving* (as practiced in computational logic), but is definitely distinct from them. It is based on what I call *the ansatz ansatz* ([12][13], see [22] for a philosophical discussion). In this methodology, one "teaches" the computer how to "conjecture an answer" to a problem, and then "teaches" that very same computer to prove its own conjectures rigorously. The novelty is that *both* the conjecturing and the proving are automatically done by the computer. This does not mean that human mathematicians are superfluous. Quite the contrary! Someone has to "teach" the computer, i.e. design algorithms and meta-algorithms for both proving and conjecturing. In my experience, this act of "teaching" the computer how to do mathematics is at least as challenging as doing mathematics "by hand", and in my humble opinion, time much better spent, since the vast potential of the computer is still very underutilized and underrated, and it is important to have mathematicians, like myself and my students, who are dedicated to that activity, that I believe will soon revolutionize mathematics.

**The Parts**

In the last five years, in collaboration with my students and other researchers, 32 research articles were written, on various topics. Most of these papers are accompanied by long and sophisticated Maple packages, that are downloadable from my website, free of charge.

I will now very briefly list the *specific* accomplishments.

**Wilf-Zeilberger Algorithmic Proof Theory** ([2][5][9][10][17]): The so-called WZ theory was further extended and applied, in collaboration with my student Moa Apagodu, who graduated in 2006.

**Combinatorial Games**([1][25]): The Theory of Combinatorial Games is a great arena for *rigorous experimental mathematics*. In [1], Xinyu Sun and I prove parts of a difficult conjecture of Aviezri Fraenkel about an extension of Wythoff's game. This game is an *impartial games*. In collaboration with my recently-graduated Ph.D. student, Thotsaporn "Aek" Thanatipanonda, we ([25]) applied the methodology of rigorous experimental mathematics to *partizan* games, whose general theory was developed in the classic book *Winning Ways*[BCG], by Berlekemp, Conway, and Guy. In particular, using a *Symbolic Finite-State method*, we proved several fifteen-year-old conjecture of Jeff Erickson[Eri].

**Enumerative Combinatorics** ([3][4][7][11][12][13][15][16][18][19] [20][23][27][28][30][31]): The theory of the *Holonomic Ansatz*[12][13], that will be described in the Project Description, has already scored two exciting triumphs. In collaboration with Manuel Kauers and Christoph Koutschan, we ([30]) proved a very challenging conjecture of Ira Gessel, that states that the number of $2n$-step walks in the square lattice from the origin back to the origin, with unit steps in the four directions (Up, Down, Left, and Right), staying in the region $\{(x,y)\,|\,y \geq 0, x+y \geq 0\}$, equals $16^n(5/6)_n(1/2)_n/((5/3)_n(2)_n)$ (where $(a)_n := a(a+1)\cdots(a+n-1)$ is the rising-factorial). This is discussed further later. Perhaps an even more dramatic application of the holonomic ansatz (this time its application to determinant-evaluation described in [13]) was my recent paper([31]), also with Kauers and Koutschan, that proved (well, "almost proved") a famous conjecture of Dave Robbins and George Andrews, that dates from the early eighties, and was popularized by Richard Stanley([Sta2][Sta3]), and also mentioned in Dave Bressoud's celebrated monograph [Bre]. It concerns the $q$-enumeration of so-called *totally symmetric plane partitions*. The straight-counting ($q = 1$) case was accomplished by John Stembridge[Ste] in 1995, but the general case remained wide open. Previously, Soichi Okada ([Oka]) reduced the problem to evaluating an "innocent-looking" determinant. In [31], Kauers, Koutschan, and I, using the methodology developed in my general article [13], reduced the statement to a completely *decidable* $q$-holonomic identity, that with sufficient computing power, and *existing* (WZ-style) algorithms due to Fréderic Chyzak[Ch], should be verifiable.

**Exposition**([6][22]): My essay on *Enumerative and Algebraic Combinatorics*[6] was well-received. Part exposition, part philosophy, and part methodology, is my essay [22].

**4. List of Publications resulting from the previous NSF award 2004-2008**

1. Xinyu Sun and Doron Zeilberger, *On Fraenkel's N-Heap Wythoff Conjecture*, Annals of Combinatorics **8** (2004). 225-238.

2. Mohamud Mohammed (now Moa Apagodu) and Doron Zeilberger, *The Markov-WZ method*, Elec J. Combinatorics **11**(2004), R53. (14 pages).

3. Doron Zeilberger, *Symbolic Moment Calculus I.: Foundations and Permutation Pattern Statis-*

*tics*, Annals of Combinatorics **8** (2004), 369-378.

4. Doron Zeilberger, *Dave Robbins's Art of Guessing*, Adv. Appl. Math. **34** (2005), 939-954.

5. Mohamud Mohammed (now Moa Apagodu) and Doron Zeilberger, *Sharp Upper Bounds for the Orders of the Recurrences Outputted by the Zeilberger and q-Zeilberger Algorithms*, J. Symbolic Computation **39** (2005), 201-207.

6. Doron Zeilberger, *Enumerative and Algebraic Combinatorics*, in: "Princeton Companion of Mathematics",T. Gowers, ed., 550-561, Princeton University Press, 2008.

7. Arthur Benjamin and Doron Zeilberger, Pythagorean Primes and Palindromic Continued Fractions, INTEGERS **5(1)** (2005), A30.

8. Andrew V. Sills and Doron Zeilberger, *Disturbing the Dyson Conjecture (in a GOOD Way)*, J. Experimental Mathematics **15** (2006), 187-191

9. Doron Zeilberger, *DECONSTRUCTING the ZEILBERGER algorithm*, J. Difference Equations and its Applications **11** (2005), 851-856.

10. Doron Zeilberger, and Moa Apagodu (formerly Mohamud Mohammed) *Multi-Variable Zeilberger and Almkvist-Zeilberger Algorithms and the Sharpening of Wilf-Zeilberger Theory* , Adv. Appl. Math. **37** (2006)(Special Regev issue), 139-152

11. Doron Zeilberger, *Automatic CountTilings*, Personal Journal of Ekhad and Zeilberger, `http://www.math.rutgers.edu/~zeilberg/pj.html`, 2006.

12. Doron Zeilberger, *The HOLONOMIC ANSATZ I. Foundations and Applications to Lattice Path Counting*, Annals of Combinatorics **11**(2007), 227-239

13. Doron Zeilberger, *The HOLONOMIC ANSATZ II. Automatic DISCOVERY(!) and PROOF(!!) of Holonomic Determinant Evaluations*, Annals of Combinatorics **11**(2007), 241-247

14. Doron Zeilberger, *Symbolic Moment Calculus II.: Why is Ramsey Theory Sooooo Eeeenormoulsy Hard?*, "Combinatorial Number Theory", B. Landman et. al, editors, in Celebration of the 70th Birthday of Ronald Graham, de Gruyter, 2007. (Co-published in INTEGERS, 7(2)(2007), A34.]

15. Shalosh B. Ekhad, Vince Vatter, and Doron Zeilberger, *A Proof of the Loehr-Warrington Amazing TEN to the Power n Conjecture*, Personal Journal of Ekhad and Zeilberger, `http://www.math.rutgers.edu/~zeilberg/pj.html`, 2006.

16. Doron Zeilberger, *Symbol Crunching with the Gambler's Ruin Problem*, in: "Tapas in Experimental Mathematics", Tewodros Amdeberhan and Victor Moll, editors, Contemporary Mathematics **457** (2008), 285-292.

17. Moa Apagodu and Doron Zeilberger, *FIVE Applications of Wilf-Zeilberger Theory to Enumeration and Combinatorics*, in: "COMPUTER ALGEBRA 2006, Latest Advances in Symbolic Algorithms" [Abramov Festschrift, dedicated to Sergey Abramov's 60th birthday], Ilias S Kotsireas and Eugene V Zima, eds., World Scientific, Aug. 2007.

18. Arvind Ayyer and Doron Zeilberger, *The Number of [Old-Time] Basketball games with Final Score n:n where the Home Team was never losing but also never ahead by more than w Points*, Electronic J. of Combinatorics **14(1)** (2007), R19 (8pp).

19. Philip Matchett Wood and Doron Zeilberger, *A Translation Method for Finding Combinatorial Bijections*, Annals of Combinatorics, to appear.

20. Arvind Ayyer and Doron Zeilberger, *Two Dimensional Directed Lattice Walks with Boundaries*, in: "Tapas in Experimental Mathematics", Tewodros Amdeberhan and Victor Moll., eds., Contemporary Mathematics **457** (2008), 1-20.

21. Tewodros Amdeberhan and Doron Zeilberger, *"Trivializing" Generalizations of Some Izergin-Korepin-Type Determinants*, Discrete Mathematics and Theoretical Computer Science **9** (2007), 203-206.

22. Doron Zeilberger, *An Enquiry Concerning Human (and Computer!) [Mathematical] Understanding*, in: C.S. Calude ,ed., "Randomness & Complexity, from Leibniz to Chaitin", World Scientific, Singapore, Oct. 2007.

23. Doron Zeilberger, *Using Rota's Umbral Calculus to Enumerate Stanley's P-Partitions*, Adv. Applied Mathematics **41**(2008), 206-217.

24. Manuel Kauers and Doron Zeilberger, *Experiments With a Positivity Preserving Operator*, Experimental Mathematics, to appear.

25. Thotsaporn "Aek" Thanatipanonda and Doron Zeilberger, *A Symbolic Finite-State Approach For Automated Proving of Theorems in Combinatorial Game Theory*, J. Difference Eq. Applications, to appear.

26. Moa Apagodu and Doron Zeilberger, *Searching For Strange Hypergeometric Identities By Sheer Brute Force*, INTEGERS **8**(2008), A36.

27. Manuel Kauers and Doron Zeilberger, *The Quasi-Holonomic Ansatz and Restricted Lattice Walks*, To appear in J. of Difference Equations and Applications [special issue in honor of Gerry Ladas' 70th Birthday].

28. William Y.C. Chen, Jing Qin, Christian M. Reidys and Doron Zeilberger, *Efficient Counting and Asymptotics of k-noncrossing Tangled Diagrams*, submitted.

29. Yuri Bahturin, Amitai Regev and Doron Zeilberger, *Commutation Relations and Vandermonde Determinants*, to appear in European J. Combinatorics.

30. Manuel Kauers, Christoph Koutschan, and Doron Zeilberger, *Proof of Ira Gessel's Lattice Path Conjecture*, submitted.

31. Manuel Kauers, Christoph Koutschan, and Doron Zeilberger, *A Proof of George Andrews' and Dave Robbins' q-TSPP Conjecture (modulo a finite amount of routine calculations)*, submitted.

32. Doron Zeilberger, *Proof of a conjecture of Philippe Di Francesco and Paul Zinn-Justin related to the qKZ equations and to Dave Robbins' two favorite combinatorial objects*, Personal Journal of Ekhad and Zeilberger, `http://www.math.rutgers.edu/~zeilberg/pj.html`, 2006.

**5.** As already mentioned above, most of my papers are accompanied by Maple packages that are available, free of charge, from my homepage `http://www.math.rutgers.edu/~zeilberg/`. In addition, there are quite a few packages that belong to forthcoming papers, or stand by themselves. Some of them are of a rather general scope, and should be useful to researchers in combinatorics, number theory, analysis, statistical physics, and possibly other areas.

**6.** A large part of the proposed research is a direct continuation of the previous research, but there are also new directions, in which the connection is less obvious.

**7. Education and Human Resources Statement**

During the discussed period (2004-2008), six students received their Ph.D. degree under my guidance. **Xinyu Sun** (2004), who was visiting assistant professor at Texas A&M for three years, working with Catherine Yan, and who is currently visiting assistant professor at Tulane University, collaborating with Experimental Mathematics guru Victor Moll; **Xiangdong Wen** (2005), who works for Wolfram Research; **Vince Vatter** (2005), who was a postdoctoral fellow at St. Andrews University for two years, was then offered a prestigious NSF posdoc at MIT, (to work with Richard Stanley), but decided, to my great disappointment, to "sell out", and accepted an offer from the hedge fund DE Shaw. To my utmost joy, he was so disgusted by the corporate culture that he resigned nine days later, and returned to academia, and has just started a three year position at Dartmouth College; **Moa Apagodu** (2006), who accepted a tenure-track assistant professorship at Virginia Commonwealth University; **Lara Pudwell** (2008), who was recruited by her undergraduate alma mater, Valparaiso University (Indiana), and is now tenure-track assistant professor there. Lara was supported, in part, by this grant, and did very impressive work on *Enumeration Schemes* for forbidden patterns in permutations. Lara is also a great educator and received a university-wide award for excellence in contributions to undergraduate education; **Thotsaporn "Aek" Thanatipanonda**(2008), who is now a visiting Assistant Professor at Dickinson College. Aek received a university-wide excellence in research award this Spring, one of six, competing with a pool of four thousand students in *all* subjects; **Arvind Ayyer**(2008), who was jointly supervised by Joel Lebowitz, and whose Ph.D. is in physics. Arvind started (Oct. 2008) a two-year postdoctoral position at the Institute of Theoretical Physics in Saclay, France.

In Spring 2008 I expect to graduate two more students. **Eric Rowland**, who works on enumeration and discrete dynamical systems, and who recently made quite a splash with his "prime-generation" algorithm, that was written-up in *Pour La Science*, and Ivars Peterson's MAA Column *The Mathematical Tourist*, as well as Jeff Shallit's widely read blog *Recursively*; **Paul Raff**, who does very interesting work on combinatorial problems with applications to epidemiology, and on counting spanning trees. In addition, I have two more current Ph.D. students: **Andrew Baxter**, and **Emilie Hogan**. Their progress is very satisfactory.

## PROPOSED RESEARCH: RIGOROUS EXPERIMENTAL COMBINATORICS

In his preface to $A = B$ ([PWZ]), Don Knuth famously wrote:

"*Science is what we understand well enough to explain to a computer. Art is everything else we do.*"

Then he went on to comment how so-called Wilf-Zeilberger Algorithmic Proof Theory turned an "important part of mathematics" (hypergeometric summation and integration) from an Art into a Science.

Important as hypergeometric summation and integration may be, they are still rather small, specialized, areas. I strongly believe that WZ theory is a harbinger of analogous algorithmic proof theories in many other branches of mathematics, and a large part of my research agenda consists in trying to discover and use them. Of course, one should be flexible, and not insist on emulating WZ theory completely. For example, if one can develop an algorithmic proof theory for a class of problems that only works *sometimes* (i.e. that is not guaranteed to *always* succeed), that would be nice too.

A close look at WZ theory reveals that it is essentially "systematic and inspired guessing", so let's pause and talk about *guessing*.

### The Art of Guessing

*Guessing*, or more "respectably", *conjecturing*, is a very important part of mathematical research, that unfortunately was suppressed for many years, until it made a comeback when computers came along. A very notable exception was George Polya's marvelous approach ([Po1][Po2]), that I believe should, and could, be adapted to the computer age (see [22]). All the great mathematicians of the past "got their hands dirty", and arrived at their beautiful theorems and conjectures after extensive experimentation (with paper-and-pencil, of course, and in Archimedes' case, sand-and-stick). I am sure that they would have done much more if they had a computer algebra system like Maple or Mathematica.

In his delightful essay on Experimental Mathematics, Herb Wilf ([Wil]) describes "regular" experimental mathematics, as it is practiced today, as follows.

**1.** Wondering, by a human, what a "particular situation looks like in detail".

**2.** Computer experimentation to show the structure of that situation for a selection of small values of the parameters.

**3.** The [human] mathematician *gazes* at the computer output, attempting to see or to codify some pattern, that hopefully leads him or her to formulate a *conjecture*.

**4.** Human-made proof of the human-made conjecture (that was computer-inspired).

### Turning Guessing from an Art into a Science

In *my* version of experimental mathematics, the first two steps remain the same, but the last two are replaced by:

**3'.** The computer (automatically) tries to discovers a pattern, following an *ansatz* (see below) specified by the human mathematician.

**4'.** The computer discovers a proof, all by itself, by using a "proof-ansatz" specified by the human, and mathematical induction, that has been programmed into the computer, thereby giving a completely rigorous proof.

So, in the present approach to mathematical research, one expects much more from the computer than just being (as George Andrews quiped) a "pencil with power-steering". Computers do the *actual* mathematics, while humans are engaged in "meta-mathematics", but not in the traditional sense of logic and set theory, but in "teaching" computers to discover and prove theorems within existing *ansatzes*, and, perhaps even more importantly, developing new ansatzes. But let's first define the *keyword* **ansatz**.

**What is an Ansatz?**

According to Eric Weisstein's `mathworld.com` wonderful website, in an entry contributed by Mark D. Carrara[Wei],

*"An ansatz is an assumed form for a mathematical statement that is not [necessarily] based on any underlying theory or principle."*

In other words, you make a *wild guess* that the desired solution has a certain *form*, featuring some *undetermined* coefficients, "plug" that form into the conditions of the problems, and try to solve for the coefficients. If *in luck*, you find a solution, and then, since *the proof of the pudding is in the eating*, you have an **a posteriori** justification for choosing that ansatz, and more importantly for your short-term goals, you have **solved** the problem! In addition, your present success will give you more confidence that this ansatz might possibly work for similar problems in the future.

**My General Research Agenda:** *Exploit Known ansatzes to tackle challenging problems in mathematics, and develop new ones.*

**Part I: Exploiting Known Ansatzes**

Even within the already known ansatzes, there is still a lot of work to be done. Some of the better known ones (see [22] for more detail) are the *polynomial ansatz*, the *C-finite* ansatz (sequences whose generating function is a rational function), the (what I call) *Schützenberger* ansatz (sequences whose generating function is an algebraic formal power series), and the *holonomic* ansatz (the framework for WZ theory). I propose to utilize each of these for various problems, that I will now describe.

**The Polynomial Ansatz**

A very striking example can be found in [Z1]. Since it is so brief, let me quote it in its entirety.

*For a permutation $\pi$, let inv $\pi$ be the number of $(i, j)$, such that $1 \leq i < j \leq n$ and $\pi[i] > \pi[j]$, and maj $\pi$ be the sum of $i$, such that $1 \leq i < n$ and $\pi[i] > \pi[i+1]$. Svante Janson asked Don Knuth, who asked me, about the covariance of inv and maj. The answer is $\binom{n}{2}/4$. To prove it, I asked Shalosh to compute the average of the quantity $(inv\ \pi - E[inv])(maj\ \pi - E[maj])$ over all permutations of a given length $n$, and it gave me, for $n = 1, 2, 3, 4, 5$, the values $0, 1/4, 3/4, 3/2, 5/2$,* *respectively. Since we know a priori*[1] *that this is a polynomial of degree* $\leq 4$*, this must be it!* $\Box$

But "brute brute force" can only go so far. If instead of trying to find the covariance, that is the average of $(inv\ \pi - E[inv])(maj\ \pi - E[maj])$, one wanted to use the above naive approach to find, the "higher-covariances"

$$C_{r,s}(n) := \frac{1}{n!} \sum_{\pi \in S_n} (inv\ \pi - E[inv])^r (maj\ \pi - E[maj])^s \quad ,$$

a polynomial of degree $\leq 3(r+s)/2$ in $n$, say for $r = 10$, and $s = 10$, one would need 31 data points, and the computer would have to perform well over 31! (roughly $0.822 \cdot 10^{34}$) additions. Not very realistic!

But why so brutish? Here is an alternative way to generate the needed 31 terms of the sequence that would enable you to compute them much more efficiently, before you fit them into a polynomial of degree 30.

Define the bi-variate generating function

$$A_n(q_1, q_2) = \sum_{\pi \in S_n} q_1^{inv \pi} q_2^{maj \pi} \quad .$$

Unlike $A_n(q,1)$ and $A_n(1,q)$, that are both given by the closed-form expression $[n]! = (1-q)(1-q^2)\cdots(1-q^n)/(1-q)^n$ (the so-called $q$-analog of $n$ factorial), both classical results that go back to Netto and MacMahon respectively, there is no known "closed-form" for $A_n(q_1, q_2)$. Nevertheless, one can form a *recurrence* scheme, that would enable one to easily compute these polynomials for, say, $n \leq 60$. It goes as follows.

Let $F_{n,i}(q_1, q_2)$ be the weight-enumerator according to the weight $q_1^{inv\pi} q_2^{maj\pi}$, of the set of permutations that end in $i$. Once we know $F_{n,i}$, we can compute $A_n$, since $A_n = \sum_{i=1}^n F_{n,i}$. A fast way to compute $F_{n,i}(q_1, q_2)$ is in terms of the recurrence (whose proof is omitted here, but is fairly straightforward)

$$F_{n,i}(q_1, q_2) = q_1 F_{n,i+1}(q_1, q_2) + q_1^{n-i}(q_2^{n-1} - 1) F_{n-1,i}(q_1, q_2) \quad ,$$

subject to the *boundary conditions* $F_{n,n}(q_1, q_2) = A_{n-1}(q_1, q_2)$, and the *initial condition* $F_{1,1}(q_1, q_2) = 1$. To get $C_{r,s}(n)$ (the higher covariances), we simply use

$$C_{r,s}(n) = \frac{1}{n!} (q_1 \frac{\partial}{\partial q_1})^r (q_2 \frac{\partial}{\partial q_2})^s \left( \frac{A_n(q_1, q_2)}{q_1^{n(n-1)/4} q_2^{n(n-1)/4}} \right) \Big|_{q_1=1, q_2=1} \quad ,$$

and now we can easily ask the computer to crank out many terms of $A_n(q_1, q_2)$, that would immediately allow us to compute the first 31 terms (say) of the sequence $\{C_{10,10}(n)\}$, that, in turn, would

[1] This is the old trick to compute moments of combinatorial 'statistics', described nicely in Graham, Knuth, and Patashnik's 'Concrete Math', section 8.2, by changing the order of summation. It applies equally well to covariance. Rather than actually carrying out the gory details, we observe that this is always a polynomial whose degree is trivial to bound.

enable us, by fitting these data into a degree-30 polynomial in $n$, to find an explicit polynomial expression. (In general, for $C_{r,s}(n)$ we would need $3(r+s)/2+1$ terms, of course).

But what about $C_{100,200}(n)$? Then even this improved approach will run out of memory. But there is yet another, more efficient way to crank out $C_{r,s}(n)$ directly. Using the above recurrence for $F_{n,i}$ it is possible to get recurrences directly for $A_{r,s}(n)$, for much lager $r$ and $s$. But who cares? Do we really want to see the fully expanded degree-450 polynomial expression for $C_{100,200}(n)$? Fortunately, this approach can do much more. It can give us just the leading term in the asymptotics, from which the human can easily guess the asymptotic expression (that the computer (or human) can then prove by induction). From this data it is not too hard to conjecture (and then prove) the following asymptotics for the *normalized generalized covariances* (or "higher-correlation coefficients"):

$$N_{2r,2s}(n) \ := \ \frac{C_{2r,2s}(n)}{C_{2,0}(n)^r C_{0,2}(n)^s} = \frac{(2r)!}{r!2^r}\frac{(2s)!}{s!2^s} + O(1/n) \quad ,$$

and $N_{2r+1,2s}(n), N_{2r,2s+1}(n), N_{2r+1,2s+1}(n)$ are all $O(1/n)$. *Now* this is of great human interest! It implies, via the so-called *method of moments*, that the random variables *maj* and *inv*, once (as usual in limit laws in probability) centralized and divided by their standard deviations, are *joint asymptotically normal*, i.e., as $n \to \infty$, closely resemble two independent standard Gaussians.

By hindsight, this theorem could have been proved by entirely human means, but it was inspired by computer, and furthermore, with a computer, we can get many more terms in the asymptotic expansion. There are numerous other natural permutation statistics, and I believe that analogous "limit theorems" can be proved for them, and it may be also possible to consider the joint-distribution of more than two at a time. This brings us to:

**Research Problem 1**: Use the Polynomial ansatz as the framework for Rigorous Experimental studies of a Statistical (symbolic) theory of Permutation Statistics.

We will meet permutation statistics again, when I will talk about them in the context of a yet-to-be-developed *multi-basic* ansatz.

**The C-finite ansatz**

A sequence $\{a(n)\}_{n=0}^{\infty}$ is $C$-finite if it satisfies a homogeneous linear recurrence equation with constant coefficients. Equivalently, if its (ordinary) generating function is a rational function of $x$, i.e. there exist polynomials $P(x)$ and $Q(x)$ such that

$$\sum_{n=0}^{\infty} a(n)x^n \ = \ \frac{P(x)}{Q(x)} \quad .$$

The simplest (non-constant) $C$-finite sequence is $\{2^n\}$ (whose generating function is $1/(1-2x)$) and the simplest one of order higher than one, is the sequence of Fibonacci numbers $\{F_n\}$ (whose generating function is $x/(1-x-x^2)$). The polynomial ansatz is a subansatz of the present one, since the generating function of a polynomial sequence of degree $d$ is a rational function with denominator $(1-x)^{d+1}$.

As simple as this ansatz is, it too can be used in rigorous experimental mathematics. An amusing example is given in [11], where my computer, Shalosh B. Ekhad *rigorously* proved the Kasteleyan-Fisher&Temperly ([K][FT]) celebrated formula for the number of ways of tiling an $m$ by $n$ rectangle with "dimers" ($2 \times 1$ and $1 \times 2$ tiles, i.e. "dominoes) for fixed $m$ (in practice $m \leq 20$) but *general*(!) $n$. This experimental approach could, presumably, with *some* human help, lead to a new proof of the general theorem.

But in a way, Kasteleyan and Fisher&Temperly were lucky. It so happened that this problem had such a structure that human techniques worked. In general such problems, that arise in statistical physics, are notoriously difficult, and physicists resort to non-rigorous and approximate heuristic methods in order to say interesting things. For example, the analogous problem of tilings a rectangle with both dimers and *monomers* ($1 \times 1$ tiles) is wide open, as are the Ising model with magnetic field in two dimensions, and the three-dimensional Ising model. A complete solution is, of course, a "long shot", but I believe that using the $C$-finite ansatz to study *finite* approximations, may pave the way to, who knows?, a new ansatz that will take care of these open problems, or at least indicate their intractability. To summarize, we have:

**Research Problem 2**: Use the $C$-finite ansatz to study "finite-forms" of seemingly intractable problems in combinatorial models that arise in statistical physics.

Another grist-for-the-mill for the "humble" $C$-finite ansatz is Ramsey Theory and Szemerédi's theorem ([Sze]). Recall that Szémeredi's theorem asserts that for any given integer $k$, and any density $\delta$, there exists an integer $N_0(k, \delta)$ such that for $n \geq N_0$ any subset of $[1, n]$ with at least $\delta n$ elements is guaranteed to have an arithmetical progression of length $k$. Szemerédi only proved that $N_0(k, \delta) < \infty$, and the implied bounds were astronomical. Gowers[Go] brought them down, and for the $k = 3$ case (Roth's theorem), Bourgain([Bou1][Bou2]) brought them even further down. But neither is believed to be sharp, and there is probably lots of room for improvement. Surprisingly, this problem can be "approximated" by sequences that belong to the $C$-finite ansatz, by looking at those arithmetical sequences with bounded difference, using an extension of the Goulden-Jackson ([GJ]) method due to Noonan and myself ([NZ]). This brings us to:

**Research Problem 3**: Use the $C$-finite ansatz to study "finite-forms" of difficult problems in Ramsey Theory and Additive Number Theory. Try to get insight both from the output, and from the method.

**The Schützenberger Ansatz**

A sequence $a(n)$ belongs to that ansatz whenever its generating function $f(x)$ is a solution of an algebraic equation whose coefficients are polynomials in $x$, i.e. there exists a polynomial in two variables $P(x, y)$ such that $P(x, f(x)) \equiv 0$. For the special case where $P(x, y)$ is of degree one in $y$ we are back to the $C$-finite case. The paradigmatic example of such a sequence is the sequence of Catalan numbers, whose generating function, $\phi(x)$, satisfies $\phi(x) = 1 + x\phi(x)^2$. See my Maple package SCHUTZENBERGER available from my website, and the very useful Maple package `gfun` developed by Bruno Salvy and Paul Zimmerman [SZ]. Such sequences arise naturally in tree-enumeration, and lattice path-counting. The *ecóle bordelaise*, under the leadership of Xavier Viennot and Mireille Bousquet-Mélou (see e.g. [BM]) developed a marvelous theory. While they do use computer algebra systems quite extensively, they do so mainly as a symbolic *calculator*, in the traditional mode. I believe that there is lots of work to be done in upgrading to my style of experimental mathematics.

One way that such formal power series arise is related to the problems in statistical physics that I have already mentioned. When a physicist is interested in a quantity, let's call it $A_{m,n}(z)$, it depends on a parameter $z$ (e.g. the temperature). It often happens that for fixed $m$, the sequence $A_{m,n}(z)$ is $C$-finite in $n$ (with increasingly complicated generating functions as $m$ gets larger), but the physicists are really only interested in the *thermodynamic limit*

$$f(z) := \lim_{m\to\infty} \lim_{n\to\infty} A_{m,n}(z)^{1/mn} \quad ,$$

and its analytical behavior as a function of $z$ (a singularity corresponds to a *phase transition*, and its mathematical nature corresponds to the physical nature of the phase transition). It is easy to see that whenever $\{A_{m,n}(z)\}_{n=0}^{\infty}$ is $C$-finite,

$$f_m(z) := \lim_{n\to\infty} A_{m,n}(z)^{1/n} \quad ,$$

belongs to the present ansatz, and if we'll understand $f_m(z)$ well enough, it may lead us to $f(z) = \lim_{m\to\infty} f_m(z)^{1/m}$, by moving up to a higher ansatz (in the case of the 2D Ising model, it happens to be the holonomic ansatz, to be discussed shortly). So progress on the Schützenberger ansatz may have applications in mathematical physics. In short, we have:

**Research Problem 4**: Extend the beautiful, but human, work of the Bordeaux school (and other researchers), to make more active use of computer-generated (rigorous) mathematics.

**My favorite Ansatz: The Holonomic Ansatz**

A sequence $a(n)$ (of a single discrete variable $n$) is called *holonomic* if it satisfies a (homogeneous) *linear recurrence equation with polynomial coefficients*, i.e. there exists a positive integer $L$ (the order), and $L+1$ polynomials $p_0(n), p_1(n), \ldots, p_L(n)$ such that

$$\sum_{i=0}^{L} p_i(n) a(n+i) = 0 \quad .$$

This concept was implicit for a long time, but was first explicated in Richard Stanley's seminal paper [Sta1]. Stanley called such sequences *P-recursive.*

In [Z2], I show that almost everything in sight in enumerative combinatorics, and a lot elsewhere, is holonomic. I also gave a "slow" algorithm that was later made much faster by Frederic Chyzak[Ch] and others, and in later developments I found much faster algorithms for important special cases. But there is an alternative, more "empirical" approach, that I described in [12] and [13], that applies it to *lattice path counting*, and *automatic determinant-evaluation*. respectively. Note that WZ-theory is not always applicable to these problems, so the scope of the present approach is different.

To illustrate the ideas, let me sketch the main strategy in my recent proof, with Manuel Kauers and Christoph Koutschan ([30]) of Ira Gessel's notorious lattice path conjecture, already mentioned above, that states that the number of $2n$-step walks in the square lattice from the origin back to the origin, with unit steps in the four directions (Up, Down, Left, and Right), staying in the region

$\{(x,y)|y \geq 0\,,\ x+y \geq 0\}$, equals $16^n (5/6)_n (1/2)_n / ((5/3)_n (2)_n)$ (where $(a)_n := a(a+1)\cdots(a+n-1)$ is the rising-factorial).

Consider the more general discrete function $A(x,y;n)$, the number of ways of getting from the origin to a general point $(x,y)$ in $n$ steps. $A(x,y;n)$ obviously satisfies:

$$A(x,y;n) = A(x-1,y;n-1)+A(x,y-1;n-1)+A(x+1,y;n-1)+A(x,y+1;n-1) \qquad (\textit{Recurrence})$$

in $\{(x,y)|x+y \geq 0, y \geq 0\}$, subject to the *initial condition* $A(x,y;0) = 0$ if $(x,y) \neq (0,0)$, and $A(0,0;0) = 1$, and the *boundary conditions* $A(x,y;n) = 0$ if $x+y < 0$ or $y < 0$. A natural approach would be to try and conjecture an "explicit" expression for this general $A(x,y;n)$, prove it by induction on $n$ using (*Recurrence*), and then plug-in $(x,y) = (0,0)$ to get Gessel's conjectured explicit expression for $A(0,0;2n)$ (obviously $A(0,0;2n+1) = 0$).

Alas, this approach, when taken literally, seems doomed to failure. As far as anyone knows, $A(x,y;n)$ can't be written in closed-form. But the *holonomic* ansatz provides an alternative, and one can just broaden the definition of "closed-form" to that of "satisfying an appropriate partial recurrence". Our computers fond a *partial recurrence* equation of the form

$$\sum_{i=0}^{L} p_i(n) A(x,y;n+i) \; + \; x \sum_{0 \leq i,j,r \leq M} q'_{i,j,r}(x,y,n) A(x+i,y+j;n+r)$$

$$+ y \sum_{0 \leq i,j,r \leq M} q''_{i,j,r}(x,y,n) A(x+i,y+j;n+r) = 0 \quad ,$$

for $L = 32$, and *some* $M$, and (huge!) polynomials $p_i(n)$ and $q'_{i,j,r}(x,y,n)$,$q''_{i,j,r}(x,y,n)$. This huge equation was first conjectured (by undetermined coefficients), and then proved rigorously (automatically) by using induction and the algebra of linear partial recurrence operators with polynomial coefficients. Now, one plugs-in $x=0, y=0$ to get

$$\sum_{i=0}^{L} p_i(n) A(0,0;n+i) = 0 \quad ,$$

for some explicit polynomial coefficients $p_i(n)$ and (as it happened) $L = 32$. The rest is routine. Just verify that Gessel's nice conjectured expression also satisfies this same recurrence, and check the first 32 initial values.[2]

I call this kind of result **contingent beauty**. There is no moral reason why Gessel's expression should be so nice, it is true **just because**.

The Gessel lattice walk is just *one* example. This brings us to

**Research Problem 5**: Use the holonomic ansatz to investigate more general lattice path counting problems. What about higher dimensions? Perhaps there is a yet-to-be-discovered ansatz.

[2] This description is an over-simplification. In practice we never exhibit the $q'_{i,j,r}$ and $q''_{i,j,r}$ only prove that they exist. Since at the end of the day they will disappear, upon substituting $x=0, y=0$, they are not needed, only the fact that they exist.

**Part II: In Search of New Exciting Ansatzes**

**A Computer-Generated Theory of Permutation Statistics Based on a Multi-Basic Generalization of the Holonomic Ansatz**

In Part I we have already encountered two *permutation statistics*, $inv$, the number of inversions, and $maj$, the major index. The subject of *permutation statistics* pioneered by Major Percy MacMahon ([Mac]) and revived by Dominique Foata and his students, is now a very active part of algebraic combinatorics. Only last year John Shareshian and Michele Wachs [SW] came up with an intriguing new theorem that was missed by earlier researchers. I believe that my computer-aided approach will complement nicely the purely human research done so far.

While the generating function according to both $inv$ and $maj$ has a nice closed form ($[n]! = (1)(1+q)\cdots(1+q+\ldots+q^{n-1})$), this is not true of the joint-weight-enumerator, for $inv$ and $maj$, already introduced in Part I

$$A_n(q_1,q_2) = \sum_{\pi \in S_n} q_1^{inv \pi} q_2^{maj \pi} \quad .$$

However, as proved in my early paper [Z5], it satisfies a certain "infinite-order recurrence". This is a whole new ansatz waiting to be explored. I believe that it would prove particularly useful in a new phase in the development of the theory of permutation statistics (there are quite a few in addition to $inv$ and $maj$), significantly enhancing the pioneering work of (just to name a few) humans Eric Babson (e.g. [BS]), Jacques Désarminien (e.g. [DW]), Sergei Elizalde (e.g. [El]), Dominique Foata (e.g. [Fo]), Adriano Garsia (e.g. [GG]), Ira Gessel (e.g. [GG]), Jeff Remmel (e.g. [Re]), Einar Steingrimsson (e.g. [BS]) and Michelle Wachs (e.g. [DW],[SW]). To sum up:

**Research Problem 6**: Develop a theory of multi-basic-holonomic sequences, and apply it to initiate a new, computer-generated phase, in the classical theory of permutation statistics.

**WZ-Theory with (arbitrarily!) many variables**

WZ-theory ([WZ]) can handle multi-sums and multi-integrals with a *fixed* number of summation and/or integration signs. Consider for example, the celebrated *Selberg integral* [Se], that states that

$$\int_0^1 \cdots \int_0^1 \prod_{i=1}^{n} t_i^x (1-t_i)^y \prod_{1 \le i < j \le n} (t_i - t_j)^{2z} dt_1 \ldots dt_n$$

$$= \prod_{j=1}^{n} \frac{(x+(j-1)z)!(y+(j-1)z)!(jz)!}{(x+y+(n+j-2)z+1)!z!} \quad . \qquad (Selberg)$$

This beautiful result received quite a few additional proofs, perhaps the most elegant one being by Aomoto([Ao]). A computer-inspired, but human-generated, WZ-style proof, was give in [WZ], but there is still no *purely* computer-generated proof. It would be interesting to extend the scope of WZ theory to integrals like Selberg's with $n$ variable, where $n$ is symbolic! This is analogous to the ascent from high-school algebra to the algebra of symmetric functions with *indefinitely* many variables $x_1, \ldots, x_n$, or equivalently, *infinitely many* variables.

So far, the Selberg integral and its relatives (the Dyson, Morris, and Macdoland's constant term identities), are all with answers that are *closed-form*. It would be interesting, for example, to investigate an integral like

$$A(z,n) = \int_0^1 \dots \int_0^1 \prod_{1 \le i < j \le n} (t_i + t_j)^{2z} dt_1 \dots dt_n$$

for general $n$. WZ theory would give you a recurrence, with respect to $z$ for each specific $n$ (in practice I doubt whether one can go above $n = 4$), but what can you say about its dependence on $n$? Perhaps there is yet-another-ansatz to be discovered. This brings us to:

**Research Problem 7**: Extend WZ theory to multi-sum and integrals with indefinitely many variables. Try to find natural, Selberg-style, integrals that do not evaluate in closed-form, yet can be given a uniform description in a yet-to-be-discovered new ansatz.

In addition to the above computer-heavy research, I would also like to work on more traditional combinatorics, as follows.

**Part III: Alternating Sign Matrices and Fully Packed Loops**

Often proving a long-standing conjecture "kills" the subject. This is certainly not the case with the Mills-Robbins-Rumsey *Alternating Sign Matrix Conjecture*. The first proof [Z3], by myself, was announced in 1993, completed in 1995, and published in 1996. Shortly after, Greg Kuperberg [Ku1] came up with an ingenious, much shorter proof, using deep results in statistical physics by Izergin and Korepin that used the ubiquitous Yang-Baxter equation. Shortly after Kuperberg's proof, I ([Z4]) extended Kuperberg's approach to prove the much stronger *refined* version. The whole saga, with lots of background material, is masterfully narrated in Dave Bressoud's ([Bre]) prize-winning monograph.

Any worries that these proofs "finished" the subject will disappear if one searches `MathSciNet`, `google scholar`, or simply `arxiv.org`. The subject is booming! There is a lot of work in enumerating symmetry classes (by Kuperberg [Ku2] and others), alternative proofs (e.g. a gorgeous one by Ilse Fischer [Fis]), and connections to other intriguing combinatorial objects, for example *fully packed loops*.

There are quite a few outstanding conjectures, but perhaps the most intriguing is one due to Razumov and Stroganov (RS]), stating an amazing explicit identity about these fully packed loops. My student Arvind Ayyer and I are working on a bijective approach, that seems promising, but needs further work. Since my space is up, I will not describe it, but refer the reviewers to [AZ], that at the time of writing (Sept. 2008) of this proposal is still in preparation, but by the time the reviewers will read this proposal, will be available both at my website and at `arxiv.org`. To sum up:

**Research Problem 8**: Use bijective methods to prove the Razumov-Stroganov conjecture about so-called Fully Packed Loops.

**Conclusion: The Medium is (a large part of) the Message** [3].

The **intellectual merit** of the present proposal should be assessed on (at least) three levels. On the 'lowest' level this research contributes to combinatorics, an important field of mathematics with

[3] This is the same conclusion as in my previous proposal, from 2003. It is even truer today.

many applications to almost every branch of science, technology and human endeavor (the World Wide Web and telephone communication, CD players and pictures from Mars would be impossible without it, to mention just a few things that come to mind). The fact that I am using computers extensively in order to do my research in combinatorics should not be held for or against it, it is just another (legitimate) tool.

On the 'middle level', by using symbolic computation on a day-to-day basis, and trying to develop new algorithms (like in WZ theory), this research contributes, both directly and indirectly, to computer algebra, which is emerging as an indispensable tool not only in mathematics, but in all of science and technology.

Finally, on the 'highest-level', this research contributes to a new outlook and *awareness* in mathematical research. Mathematical research, until now, was *paper and pencil* and *a priori*, and people like Appel and Haken and Thomas Hales have to be apologetic and defensive about using computers. The computer is a mighty tool, go forth and use it! But we *humans* must think of *creative* ways of utilizing its immense potential, over-and-above its obvious use as a 'numerical and symbolic calculator' and 'brute force number- (and symbol-) cruncher'. We urgently need to develop new *methodologies* to enable us to make full use of computers. The potential applications are unforeseeable, but I am sure that future computer mathematics will make all past and present mathematics look like Mickey-Mouse stuff. But these new advances will not come by themselves. The role of the human mathematician would have to change from that of 'athlete' to that of 'coach', and this would also necessitate a change in mentality. I hope that my preaching (in my papers and the opinion column of my website), courses and seminars on Experimental Mathematics, and especially research (both the research papers viewed as case studies, and the more philosophical and methodological papers), will form a modest, yet strictly positive, beginning. **If we build it** (a new experimental methodology for mathematics), **they will come** (present and future mathematicians will practice it.)

Similarly, the **broader impact** should also be judged on more than one level. Combinatorics per se, and Computer Algebra, are both essential to science, technology, and even entertainment. But, more generally, *mathematics*, as a whole, is one of the greatest pillars of our civilization and culture, both spiritually and materially. Helping change the way we practice mathematics (for the better, I am sure), would have the broadest impact on mathematics itself. **And what's good for mathematics is good for humanity**.