How Adam Marcus and Gabor Tardos Divided and Conquered the Stanley-Wilf Conjecture (An Étude in Paramathematics)

Doron ZEILBERGER [1]

Dedicated to Apostolos Doxiadis, the First (and Foremost!) Paramathematician

...Thanks Tom[Morely] for this lovely introduction. Now I have to reciprocate and tell you that when I met Tom's thesis advisor, Dick Duffin, for the first time, in 1976, he told me, very excitedly, about his "*brilliant student Tom Morely, who is the best student he has ever had*". At the time I was only mildly impressed, OK, he is best amongst a random group of five or ten mathematicians. Only later did I find out that the set of Duffin students contains Raul Bott and Hans Weinberger...

Good afternoon! Thank you all for coming. You are really lucky, since I don't accept all invitations to speak. For most people, the **number of distinct talks** is less than the **number of talks**. It follows, by the **pigeon-hole principle** (that will show up again and again during this talk), that people don't mind **repeating themselves**. Myself, I never repeat a talk, and hence can only afford to accept so many invitations.

I came here for two reasons. First, I couldn't say no to my good friend and collaborator, Stavros Garoufalidis, and second, it gives me an opportunity to talk about a wonderful recent result in combinatorics, co-authored by one of you!, who, besides. is a *mere* (mathematical) epsilon, a first-year grad student. You are really lucky to have amongst you Adam Marcus, the co-prover, with Gabor Tardos, of the celebrated Füredi-Hajnal and Stanley-Wilf conjectures. When I suggested the title, my old friend Dick Duke wrote me E-mail that Adam already spoke about this in the combinatorics seminar a few weeks earlier. I replied that such a beautiful result can and should be repeated as often as possible.

Besides, I also had another reason for speaking about this. I wanted to give an *example* of an **interesting** mathematical talk. Most math talks are extremely boring, since most mathematicians have no clue on how to present their (or other people's) results in an interesting way. Of course, *doing* mathematics is very exciting, but somehow mathematicians, when they write it up, or lecture about it, squeeze most of the excitement out. Also they use way toooo much notation, definitions, formulas, etc. etc. and one gets an overflow of information. What they hardly ever deliver is a **good story**.

How many people have read *Uncle Petros and the Goldbach Conjecture* by Apostolos Doxiadis? [more than half of the about forty-five people raise their hands]. Now, this same Doxiadis preaches

[1] Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019, USA. `zeilberg at math dot rutgers dot edu` , `http://www.math.rutgers.edu/~zeilberg/`.
First version: Oct. 15, 2004. Transcript of a Colloq. talk given at Georgia Tech, Oct. 1, 2004, 4:05-4:55pm. Supported in part by the NSF.

that we should use **stories** when we teach math to our students. He even coined a new name for such an activity: **paramathematics**, and **paramathematicians** for its practitioners. He needed to create a new **para-profession**, since he does not believe that professional mathematicians can meet this challenge. I should also mention my former colleague John Allen Paulos, who wrote a lovely book '*Once Upon a Number*' linking math to stories.

In this talk, I want to *prove by example* that it is possible to present *cutting-edge* research in the form of a *good story*. I admit that not all of mathematics is so easily transferable to the story-medium, but combinatorics is! Alas, many combinatorialists have a complex about their field, that for a long time was scorned by the main-stream mafia, *because* the problems can be understood by everyone. (Of course, solving the problems is another story. As we all know, easily-stated problems are the hardest to prove! Witness the Goldbach Conjecture.) Because of that complex, many combinatorialists are more Catholic than the Pope and more Bourbakists than Dieudonné, and present their results in the formal, boring, style of the ruling mandarins. But, now that combinatorics is socially respectable, it is time to free ourselves from these hangups and present our beautiful combinatorial results in the most attractive form, that is through *stories*.

Let me first tell you about the **significance** of the Stanley-Wilf conjecture. I am sure that you all know that Richard Stanley is the **numero uno** guru in enumerative and algebraic combinatorics, and Herb Wilf is definitely amongst the **top five**. Yet neither Stanley, nor Wilf, nor any of the many skilled enumerators that tried very hard to prove it (yours truly included!), succeeded. What made it even more amazing and frustrating was the fact that the Marcus-Tardos proof may be presented in **half a page**. [Here I distributed a half-page handout, a copy of my short article from my Personal Journal]. Of course, it is **gorgeous**!

What is the Stanley-Wilf conjecture? Suppose that you have n people, all of different heights, and you want to arrange them in a line, in how many ways can you do this? $n!$ of course, where $n!$ is short for one times two times three ... times n. Now suppose that you want to avoid the 'pattern' abc, i.e. you forbid that there exist three people, anywhere, such that the tallest will be ahead of the second-tallest, and the second-tallest ahead of the shortest. For example, calling the 5 people $1, 2, 3, 4$ and 5, 32154 avoids abc but 15324 does not. It is well-known that the number of abc-avoiding permutations of length n is the Catalan number which is two n choose n divided by $n + 1$ which is roughly 4 to the power n, and hence of **exponential** growth, which is qualitatively less than the rate of growth of n factorial, that, thanks to Stirling, is **super-exponential**. Now the Stanley-Wilf conjecture asserts that this **exponential growth** is true for any **pattern**, of *arbitrary* length. Initially it was only known for few special cases, and later, thanks to Miklos Bona, for a rather large infinite family, but the full conjecture was wide open.

Adam Marcus and Gabor Tardos did not prove this conjecture directly, but via another conjecture, due to Zoltán Füredi and Péter Hajnal, that was previously known, thanks to Martin Klazar, to imply the Stanley-Wilf conjecture. If I only talked about their proof, I'd be done in twenty minutes. So let me first tell you, in **story-format**, the classical proofs of the celebrated Ramsey and van der Waerden theorems. Both of these warm-ups contain the same central ideas as the Marcus-Tardos proof, namely **pigeon-hole** and **modularity** (or 'divide-and-conquer'), and every combinatorialist is familiar with these proofs, and hence with these ideas. This makes it all the more frustrating that we all missed this gem.

Ramsey

Suppose that you want to make a party, where everyone has strong feelings, every pair of two people either love each other or hate each other. Note that there is no unrequited love. We want to avoid *both* **love-triangles** and **hate-triangles** (i.e. three people who mutually love [hate] each other). In a party of five, it is possible if

Abe loves Bob, Bob loves Charles, Charles loves Doron, Doron loves Ed, and Ed loves Abe,

while

Abe hates Charles, Charles hates Ed, Ed hates Bob, Bob hates Doron and Doron hates Abe .

But if the party-size is six, then Frank Plumpton Ramsey tells us that **triangles** are **unavoidable**. This fact is denoted by

$$R(3,3) = 6 \quad .$$

I will prove to you something *weaker*:

$$R(3,3) \leq 16 \quad ,$$

by describing an **algorithm** for finding a love-triangle or hate-triangle in any party of 16 people.

Call a person at a party a **lover** if he has at least as many friends as enemies, amongst the guests of the party, otherwise call him a **hater**.

Look at the shortest person, let's call him Abe. If Abe is a lover, kick out of the party all his enemies, otherwise, kick out all his lovers. Now Abe has **consistent** feelings towards all the remaining (at least) eight guests. Let Abe remain in the room, but don't consider him any longer part of the party.

Look at the shortest person amongst the remaining eight guests, let's call him Bob. If Bob is a lover, kick out of the party all his enemies, otherwise, kick out all his lovers. Now Bob has **consistent** feelings towards all the remaining (at least) four guests. Let Bob remain in the room, but don't consider him any longer part of the party.

Look at the shortest person amongst the remaining four guests, let's call him Charles. If Charles is a lover, kick out of the party all his enemies, otherwise, kick out all his lovers. Now Charles has **consistent** feelings towards all the remaining (at least) two guests. Let Charles remain in the room, but don't consider him any longer part of the party. Call the two remaining guests Doron and Ed. (If it so happens that more than two guests survived, pick any two at random to be Doron and Ed.) If Doron and Ed love each other they are both lovers, otherwise they are both haters.

At the end of the day, there are **five** surviving people in the room: Abe, Bob, Charles, Doron, and Ed, each of them with **consistent** feelings towards those that come latter in the list. By the **pigeon-hole principle**, either there are **at least three** lovers, or **at least three** haters. In the former case they form a **love triangle**, and in the latter case, a **hate triangle**. **Quod Erat Demonstatum**.

This argument can be easily extended to show that any party with 2^{2n-2} guests must contain either an n-clique or an n-anti-clique. Hence

$$R(n,n) \leq 2^{2n-2} \quad ,$$

which is not as good as the *best known* upper bound, but is pretty-damn-close, since it gives the same *asymptotics*, which is $O(4^n)$. It would be major breakthrough to replace the 4 by 3.99999999999.

van der Waerden

In this part of the talk, I will use **stories** literally, by talking about stories and their constituent elements: *letters*, *words*, *sentences*, *paragraphs*, etc.

The classical van der Waerden theorem is usually phrased in terms of **colorings** of integers. But a more up-do-date, and *apt*, formulation is via **ESL**, which are short for Equally Spaced Letters.

You must have heard, a few years ago, the big controversy about the **Bible Codes** that was espoused in Michael Drosnin's best-seller, and 'discovered' by Jerusalem mathematician Eliyahu Rips, and his two cronies. They claimed to have found amazing predictions in the Bible, by finding relevant words or names in the Torah (written one-dimensionally without any spaces) in **Equally-Spaced** places, for example at the 3003-th 4002-th, 5001-th places. The van der Waerden theorem tells you, essentially, that if the text is long enough, some 'meaningful' ESLs (of any length!) are **inevitable**.

Fix your alphabet, and let it have k letters; k could be 2 (the 0-1 alphabet of cyberworld), 4 (the ACTG alphabet of life), 22 (the aleph-tav Hebrew alphabet), 24 (the $\alpha - \omega$ Greek alphabet), 26 (the A-Z English alphabet), etc., it does not matter. Also let L be another positive integer. The van der Waerden theorem tells you that there exists a **word-length** such that **every word** of that length is **guaranteed** to have **at least one ESL** of length L of the form $aaa \dots a$, (a repeated L times) for **some** letter of the alphabet a. In the case of English, and $L = 5$ you can find a word-length N such that each and *every* one of the 26^N words of length N, is guaranteed to have an ESL of the form AAAAA, or BBBBB, or CCCCC, or $\dots$, or ZZZZZ. In other words, for every conceivable word of that length, you can find L **equally-spaced** locations that contain the **same** letter. From now on we will take an L-ESL to mean an **arithmetical progression** of locations that contain the **same letter**. For example **adoronod** contains a 3-ESL (the letter **o** at locations 3,5,7).

Reading a story (or any text for that matter) can be done on several levels. Beginning readers combine **letters** and read each letter separately. Then we get better, and learn to recognize

words, and don't think of the word *word* as 'double-u oh ar dee ' but as a single **module**. Latter on we can recognize and **encapsulate** phrases like *thank you very much*, *if you know what I mean*, and *s'il vous plait*. Some people can perhaps even perceive whole sentences at a glance, and if we were more talented we could have gone for ever.

The problem with natural language is that words have different lengths, that's why we need an extra character, the *space* in order to delimit the individual words. Also some sentences are longer than others, that's why we need a *period*. Some paragraphs are longer than others, that's why we need to start them in a new (indented) *line*, some chapters are longer than others, hence we need to use a *chapter-heading*, and so on.

Let's imagine an ideal (and artificial) language, where all text is one-dimensional, all words have the **same** length, and every conceivable string of that length is a legitimate word in the dictionary. Also decree that all sentences have the **same** number of words, all paragraphs have the **same** number of sentences etc. If that is the case, we no longer need delimiters like space, period, line-break, etc. We only need to know the word-length, sentence-length, paragraph-length etc.

Now consider a super-duper long *text* written in our alphabet. Of course, we can read it **letter-by-letter**. But we can also read it **word-by-word**, thinking of what used to be the 'vocabulary' as the new alphabet. Continuing, we can also read it **sentence-by-sentence**, **paragraph-by-paragraph**, or even **chapter-by-chapter**, with increasingly gigantic 'alphabets', in such meta-readings.

Suppose I **know** how to find, for *any* alphabet-size K, an appropriate $(L-1)$-ESL-guaranteeing word-length. I will now tell you how to use this previous **knowledge** to define a (much larger!) L-ESL-guaranteeing word-length for the k-letter original alphabet.

For the sake of definiteness let's take $k = 4$ and think of the modules as letters, words, sentences, and paragraphs.

Decree that the length of every *word* in our new language (but still with the same alphabet of size k) should be **twice** the above $(L-1)$-ESL-guaranteeing length. Then it is true that even the *first half* of every word has an $(L-1)$-ESL.

Now look at the vocabulary of all possible words. We **know** that there is a sentence-size such that every sentence must have an $(L-1)$-ESL of **words**. Decree that the length of every sentence should be **twice** that length, so that the *first half* of every sentence must have an $(L-1)$-ESL of *words*.

Now look at the 'sentence-book' of all possible sentences. We know that there is a paragraph-size such that every paragraph must have an $(L-1)$-ESL of **sentences**. Decree that the length of every paragraph should be **twice** that length, so that the *first half* of every paragraph must have an $(L-1)$-ESL of *sentences*.

Finally look at the 'paragraph-book' of all possible paragraphs. We know that there is a chapter-size such that every chapter must have an $(L-1)$-ESL of **paragraphs**. Decree that the length of every chapter should be **twice** that length, so that the *first half* of every chapter must have an $(L-1)$-ESL of *paragraphs*.

So let's consider such a gigantic **chapter**. I claim, that viewed as a *word*, in the original alphabet, it must have an L-ESL.

First consider this chapter as a *sequence of paragraphs*. We know that its *first half* contains an $(L-1)$-ESL of **identical paragraphs**, let's call these very special paragraphs (that are all **identical** to each other, and **equally spaced**)

$$P_1 \quad , \quad P_2 \quad , \quad \dots \quad , \quad P_{L-1} \quad .$$

Since this takes place in the first half, it can be continued to a would-be L-ESL

$$P_1 \quad , \quad P_2 \quad , \quad \dots \quad , \quad P_{L-1} \quad , \quad P_L \quad ,$$

but we know **nothing whatsoever** about P_L.

Now take P_1 (which is the same as P_2, and the same as P_3 , ... and the same as P_{L-1}) and **read** it as a sequence of **sentences**. We know that the **first half** of P_1 (and of course its buddies up to P_{L-1}) has an $(L-1)$-ESL of **sentences**, let's call them

$$P_{1,1} \quad , \quad P_{1,2} \quad , \dots \quad , \quad P_{1,L-1} \quad ,$$

and of course we have it all mirrored in

$$P_{2,1} \quad , \quad P_{2,2} \quad , \quad \dots \quad , \quad P_{2,L-1} \quad ,$$

$$\dots\dots\dots$$

all the way up to

$$P_{L-1,1} \quad , \quad P_{L-1,2} \quad , \quad \dots \quad , \quad P_{L-1,L-1} \quad .$$

Once again we extend it to a would-be ELS of size L

$$P_{1,1} \quad , \quad P_{1,2} \quad , \quad \dots \quad , \quad P_{1,L-1} \quad , \quad P_{1,L} \quad ,$$

except that we know **nothing** about $P_{1,L}$, and $P_{2,L}$, ... , $P_{L-1,L}$. But we do **know** that these 'random' sentences

$$\{P_{1,L} \, , \, P_{2,L} \, , \, \dots \, , \, P_{L-1,L}\}$$

are all **equal to each other**, since they are at **identical** places within their respective (**identical**) paragraphs.

Now read $P_{1,1}$ (which is an identical twin of $P_{i,j}$ for all i, j between 1 and $L-1$), as a sequence of **words**. Once again we can find an $(L-1)$-ESL of **words**

$$P_{1,1,1} \quad , \quad P_{1,1,2} \quad , \quad \dots \quad , \quad P_{1,1,L-1} \quad ,$$

that is completely identical to

$$P_{i,j,1} \quad , \quad P_{i,j,2} \quad , \quad \dots \quad , \quad P_{i,j,L-1} \quad ,$$

for all i, j between 1 and $L-1$, and extend it to a would-be-L-ESL

$$P_{1,1,1} \quad , \quad P_{1,1,2} \quad , \quad \dots \quad , \quad P_{1,1,L-1} \quad , \quad P_{1,1,L} \quad .$$

Finally, we read the word $P_{1,1,1}$ (that is completely identical with $P_{i,j,k}$, for i, j, k between 1 and $L-1$) as a sequence of **letters**. We know that the *first half* has an $(L-1)$-ESL of (individual, genuine!) **letters** (in the original alphabet), let's call them

$$P_{1,1,1,1} \quad , \quad P_{1,1,1,2} \quad , \quad \dots \quad , \quad P_{1,1,1,L-1} \quad ,$$

that once again can be extended to

$$P_{1,1,1,1} \quad , \quad P_{1,1,1,2} \quad , \quad \dots \quad , \quad P_{1,1,1,L-1} \quad , \quad P_{1,1,1,L} \quad .$$

where we know nothing about the **letter** $P_{1,1,1,L}$ (but we do know that it is the same as $P_{i,j,k,L}$ for all i, j, k between 1 and $L-1$).

Now consider the **five letters**

$$P_{1,1,1,1} \quad , \quad P_{1,1,1,L} \quad , \quad P_{1,1,L,L} \quad , \quad P_{1,L,L,L} \quad , \quad P_{L,L,L,L} \quad .$$

These are now individual letters (literally!) in the original four-letter alphabet. By the **pigeon-hole principle** (at least) two of these are the same. Say that the first and the last are identical i.e. $P_{1,1,1,1} = P_{L,L,L,L}$, then

$$P_{1,1,1,1} \quad , \quad P_{2,2,2,2} \quad , \quad \dots \quad , \quad P_{L,L,L,L} \quad ,$$

is an L-ESL of **letters**. Obviously they are equally-spaced and the first $L-1$ of them are identical by construction, and the last one is the same as the first (by assumption), hence by **transitivity of equality**, they are all **equal to each other**.

If the second and the fourth are the same, then

$$P_{1,1,1,L} \quad , \quad P_{1,2,2,L} \quad , \quad \dots \quad , \quad P_{1,L,L,L} \quad ,$$

is an L-ESL. Again the first $L-1$ of them are identical to **each other** because they are situated at identical places in identical words, and the last is identical to the first (by assumption). We can do this for each of the five choose two possibilities (in general $k+1$ choose 2). Of-course the same argument works for any size-alphabet, just continue to books, super-books, ... up to level-k super-letters.

The L-ESL-guaranteeing word-length (that turned out to be the number of letters in a chapter), even for a 2-letter alphabet, constructed by this argument (that I adapted from Khinchin's lovely little gem of a book, where he attributes it to a lady mathematician called M. A. Lukomskaya) is **eeeenormous** (as Graham/Rothschild/Spencer would put it). It was considerably brought down in 1987 by Saharon Shelah, and reduced even further, by Fields-medalist Tim Gowers, to the pitiful:

two to the power two to the power two to the power two to the power two to the power L

which is still (probably) far from the truth, that may be a **constant times two to the power L**.

Feature Presenetation: the Marcus-Tardos Proof of Füredi-Hajnal

There are n men and n women, all of different handsomeness and beauty respectively, and m (heterosexual) *love affairs* ($0 \leq m \leq n^2$). Fix k and a permutation π of size k. We want to avoid the *love-pattern* π, i.e. we forbid that there exist k men and k women such that, for $i = 1, \ldots, k$, the i-th most handsome man (amongst these k men) loves the $\pi(i)$-th prettiest woman (amongst these k women). **Zoltán Füredi** and **Péter Hajnal** conjectured that $m \leq C_k \cdot n$, for some constant C_k. Here is the **Marcus-Tardos** brilliant proof, that only uses **modularity** and **pigeon-hole** (two times!).

Fix a love-pattern π of size k, and let $f(n)$ be the maximum number of love affairs for which you can still avoid π. We want to find a *linear* upper bound for $f(n)$ (of course, trivially $f(n) \leq n^2$). Let n be divisible by a.

The first brilliant idea was to partition the men and the women into n/a clubs where the a most handsome men belong to the first club, the next a most handsome men belong to the second club, etc. and ditto for the women. A club *Loves* a club of the opposite sex if there is at least one love affair between their members.

In order to find an upper bound for $f(n)$, the total number of love affairs between individuals, all we need to know an **upper bound** for the number of *Loving clubs-pairs*, and an upper bound for the number of individual love-affairs possible between any **one particlar** man's club, and any one **one particlar** women's club (that it Loves). There are $\leq f(n/a)$ pairs of Loving club-pairs (or else a 'meta-pattern' would entail a pattern), and the number of love-affairs possible, just between members of any two specific Loving clubs is, trivially $\leq a^2$. So we get the **divide-and-conquer** inequality

$$f(n) \leq a^2 f(n/a) \quad ,$$

which immediately yields the **upper bound** $f(n) \leq n^2$. Correct, but **not very impressive**.

Can we do any better? Well, the above crude analysis is analogous to trying to find an upper bound for the total personal property in the US. We first find an upper bound for the number of people in the US, say $2.5 \cdot 10^8$, and then we find that the richest person, Bill Gates, has less than $\$4 \cdot 10^{10}$, and we get that the total personal property is at least

$$(2.5 \cdot 10^8) \times (\$4 \cdot 10^{10}) \ \leq \ \$10^{19} \quad ,$$

a true, but **far from sharp**, upper bound.

We can get a more realistic upper bound by defining a **millionaire** as someone who has more than a million dollars, and look up (in google, say) how many millionaires there are in the US. According to a quiz in the `redif.com` site, there are $2.27 \cdot 10^6$ of them. Now the number of non-millionaires is very close to the total population, so we can still use the same number, and each non-millionaire has less than $\$10^6$. As for the millionaires, we use the Bill Gates upper bound, and we now get that the total personal property in the USA is at least

$$(2.5 \cdot 10^8) \times \$10^6 + (2.27 \cdot 10^6) \times (\$4 \cdot 10^{10}) \leq \$10^{17} \quad .$$

Of course, we can find an even sharper upper bound by also bounding the number of **billionaires**, but let's stop here.

Going back to pairs of Loving clubs, it is true that they can have up to a^2 love-affairs amongst their members, but can they be too many **rich-in-love-affairs** club-pairs?

The second **brilliant idea** of Adam Marcus and Gabor Tardos was to define the analog of **millionaire** for **pairs of Loving clubs**.

A club *Adores* a club of the opposite sex if there are at least k members of the former club who love someone in the latter club.

Note that unlike 'Love', 'Adore' is not symmetric, since it may very well be that, e.g., k men of the Adoring club all have a crush on the same woman in the Adored women's-club.

It turns out that any one given club can't afford to Adore too many clubs of the opposite sex, without the **forbidden love-pattern popping up**, for the following reason.

Whenever a men's club Adores a women's club it is thanks to a certain set of k men. Now any such set of k men can't be responsible for Adoring more than $k-1$ different women's-clubs, since otherwise **all love patterns** will show up, in particular the **forbidden** one! Similarly for womens clubs Adoring men's clubs. Hence (by **pigeonhole!**), there are $\leq k\binom{a}{k} \cdot 2\frac{n}{a}$ Adoring Pairs (of clubs, from either end).

So there can't be too many Adoring club-pairs. What about a **Loving but non-Adoring** pair of clubs? Adam and Gabor claim that they are **poor-in-love-affairs**, in fact the number of love affairs between their members is $\leq (k-1)^2$, by **pigeonhole** once again! Indeed, there can be at most $k-1$ men who love *anyone* in the other club (or else it would be Adoring), and each of these men can love at most $k-1$ women in the Loved women's club (or else the men's club will be Adored by the women's club). So altogether, there can be at most $(k-1)^2$ love affairs between these two Loving-but-non-Adoring club-pair.

Now we have the **more refined divide-and-conquer** inequality

$$f(n) \leq f(\frac{n}{a}) \cdot (k-1)^2 + k\binom{a}{k}\frac{2n}{a} \cdot a^2 \quad ,$$

which simplifies to

$$f(n) \le (k-1)^2 \cdot f(\frac{n}{a}) + 2ka\binom{a}{k} \cdot n \quad ,$$

that immediately entails

$$f(n) \le \frac{2a^2k\binom{a}{k}}{a-(k-1)^2} \cdot n \quad .$$

Now take any $a \ge (k-1)^2+1$. Marcus and Tardos took $a = k^2$, getting $C_k = 2k^4\binom{k^2}{k}$. $\square$

Martin Klazar showed, in 2000, that **Füredi-Hajnal** implies **Stanley-Wilf**. It will take me another two minutes to describe it, but I hate to go overtime, so just look at the handout I gave you, or search google for **"Marcus-Tardos"** and you will see it right away. Thanks so much!