

Four Proofs of the Ballot Theorem ¹

Marc Renault
Shippensburg University
Shippensburg, PA 17257
MSRenault@ship.edu

Introduction

One of the great pleasures in mathematics occurs when one considers several different proofs of a single result. In fact, when one considers the myriad proofs of the Pythagorean theorem and the irrationality of $\sqrt{2}$ constructed over the centuries, it seems we humans can never be satisfied with just one proof. Why do we continue to devise new approaches to known results? There is something in the reasoning itself that brings insight to the problem beyond what the result tells us, like looking at a sculpture from many different perspectives to appreciate it as fully as possible.

In this article we present four proofs of the ballot theorem, describe some of the history surrounding each of the proofs, and consider the different perspectives that each brings to the problem.

The Ballot Problem: Suppose that in an election, candidate A receives a votes and candidate B receives b votes, where $a \geq kb$ for some positive integer k . Compute the number of ways the ballots can be ordered so that A maintains more than k times as many votes as B throughout the counting of the ballots.

¹This paper has been accepted by MATHEMATICS MAGAZINE, April 2007.

The Ballot Theorem: The solution to the ballot problem is $\frac{a - kb}{a + b} \binom{a + b}{a}$.

Let us call a permutation of the ballots *good* if A stays ahead of B by more than a factor of k throughout the counting of the ballots, and *bad* otherwise. Since the total number of distinct permutations of the $a + b$ ballots is

$$\binom{a + b}{a} = \frac{(a + b)!}{a!b!},$$

the theorem tells us that if all ballot permutations are equally likely, then the probability of a good permutation occurring is $(a - kb)/(a + b)$.

In 1887 Joseph Bertrand [8] introduced the ballot problem for the case $k = 1$, gave its solution, outlined an inductive proof, and asked if a “direct solution” could be found. Almost immediately after Bertrand posed his question, Émile Barbier [5] stated and provided a solution to the ballot problem for arbitrary k , but without any proof. Very shortly after Barbier, Désiré André [4] produced a short combinatorial proof of the ballot theorem for $k = 1$. In 1923 Aeppli [2] announced that he had the first proof of the ballot theorem for $k \geq 1$, and he directed interested readers to see his Ph.D. thesis [3, pp. 11-15] for the proof. Takács [30] supplies a nice account of the historical development of various ballot theorems, and gives several proofs of the ballot theorem, including the original proofs by André and Aeppli.

Proof 1: Count the Bad Ballot Permutations

André’s approach for the case $k = 1$ is to count the number of bad ballot permutations and subtract that from the number of all ballot permutations to obtain the number of good ballot permutations. Briefly, André supposes

that a ballots are marked “ A ” and b ballots are marked “ B ”. He first notes that every ballot permutation starting with B is bad, and there are $\binom{a+b-1}{a}$ of these. Through a reversible procedure, he demonstrates a one-to-one correspondence between the bad ballot permutations starting with A , and all permutations consisting of a A ’s and $b-1$ B ’s. Again, these number $\binom{a+b-1}{a}$. He concludes that the number of bad ballot permutations is $2\binom{a+b-1}{a}$, and the ballot theorem then follows by simplifying $\binom{a+b}{a} - 2\binom{a+b-1}{a} = \frac{a-b}{a+b}\binom{a+b}{a}$.

The ballot problem and its solution caught the imagination of mathematicians, and many variations of André’s proof have appeared throughout the years. For instance, Percy MacMahon [18] applied his deep theory of partitions to the problem. The most famous and elegant of these variations is the “reflection method” (often misattributed to André) in which ballot permutations are represented as lattice paths and portions of the bad paths are reflected across a line. This method was developed in the pair of papers [1] and [19] in 1923. Interestingly, the reflection method fails to generalize in a way that solves the ballot problem for $k > 1$. See [23] for more detail on André’s original proof, the reflection method, and extending André’s original proof to the case $k \geq 1$.

In 2003 Goulden and Serrano [14] provided a clever proof of the ballot theorem (for $k \geq 1$) using André’s “count the bad ballot permutations” approach [14], and we present a variation of that proof here. Their proof *rotates* a portion of a lattice path instead of reflecting it.

Proof 1. We can think of a ballot permutation as a lattice path starting at $(0, 0)$ where votes for A are expressed as upsteps $(1, 1)$ and votes for B are expressed as downsteps $(1, -k)$. We seek the number of such paths with a

upsteps and b downsteps where no step ends on or below the x -axis. Paths that remain above the x -axis (after the origin) are *good*, while those with steps that end on or below the x -axis are *bad*. A downstep that starts above the x -axis and ends on or below it is called a *bad step*.

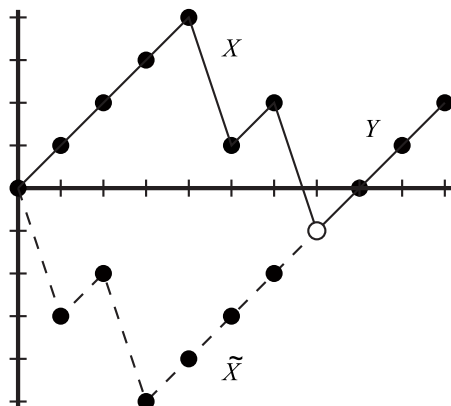


Figure 1: Example with $k = 3$. $XY \in \mathcal{B}_1$ and $\tilde{X}Y \in \mathcal{B}_3$.

For $0 \leq i \leq k$, let \mathcal{B}_i denote the set of bad paths whose first bad step ends i units below the x -axis. Clearly these $k + 1$ sets are disjoint and their union is the set of *all* bad paths. Notice that the paths in \mathcal{B}_k are exactly those paths that start with a downstep, and so $|\mathcal{B}_k| = \binom{a+b-1}{a}$. We now show that for any $i \neq k$ we actually have $|\mathcal{B}_i| = |\mathcal{B}_k|$.

Let P be a path in \mathcal{B}_i ($i \neq k$), and identify the first step of P that ends i units below the x -axis. Let X be the initial segment of P that ends with that step and write $P = XY$. Let \tilde{X} denote the path that results from rotating X by 180° , exchanging its endpoints; see Figure 1. Since X ends with a downstep, \tilde{X} starts with a downstep, and consequently $\tilde{X}Y \in \mathcal{B}_k$.

The same process converts a path in \mathcal{B}_k into a path in \mathcal{B}_i ($i \neq k$). If

$P \in \mathcal{B}_k$, then identify the first step that ends i units below the x -axis. Let X denote the initial segment of P that ends with that step and write $P = XY$. Since X necessarily ends with an upstep, we have $\tilde{X}Y \in \mathcal{B}_i$.

Thus each of the $k + 1$ sets \mathcal{B}_i have cardinality $\binom{a+b-1}{a}$, and the number of good paths is

$$\binom{a+b}{a} - (k+1)\binom{a+b-1}{a} = \frac{a-kb}{a+b}\binom{a+b}{a}. \quad \square$$

Of particular interest in this proof is the fact that the sets \mathcal{B}_i of bad ballot permutations all have the same cardinality, regardless of i . We say that these sets *uniformly partition* the set of bad paths. In Proof 4 we will see another instance of a uniform partition.

As often happens in mathematics, it appears that the above 2003 proof is essentially a rediscovery of the 1923 proof by Aeppli. Aeppli's proof of the ballot theorem appeared in his dissertation [3], and was not widely available until Takács [30] provided a "somewhat modified version" of the proof in 1997. In his proof, Aeppli uses no geometric reasoning, and instead of counting the number of good ballot permutations he computes the probability that a ballot permutation is good (provided, of course, that all ballot permutations are equally likely). He partitions the bad ballot permutations in exactly the same manner as the preceding proof does; moreover, to show a one-to-one correspondence he reverses an initial portion of a ballot permutation, which is geometrically equivalent to rotating an initial portion of a lattice path.

Proof 2: Induction

In 1887 Barbier stated the ballot theorem for $k \geq 1$ without proof. If he had a proof, one supposes it followed the inductive proof that Bertrand sketched for the case $k = 1$. An inductive proof is not difficult to construct, and no record seems to exist for the “first” such proof of the ballot theorem. The proof below is similar to that found in Takács [30].

Proof 2. Let $N_k(a, b)$ denote the number of ways the $a + b$ ballots ($a \geq kb$) can be ordered so that candidate A maintains more than k times as many votes as B throughout the counting of the ballots. The conditions $N_k(a, 0) = 1$ for all $a > 0$ and $N_k(kb, b) = 0$ for all $b > 0$, are easily verified by considering the statement of the ballot problem, and they both satisfy $N_k(a, b) = \frac{a-kb}{a+b} \binom{a+b}{a}$.

For $b > 0$ and $a > kb$, we see that $N_k(a, b) = N_k(a, b-1) + N_k(a-1, b)$ by considering the last vote in a ballot permutation. By induction, this quantity is $\frac{a-k(b-1)}{a+b-1} \binom{a+b-1}{a} + \frac{a-1-kb}{a+b-1} \binom{a+b-1}{a-1}$ which simplifies to $\frac{a-kb}{a+b} \binom{a+b}{a}$ as needed. \square

Proof 3: The Cycle Lemma

In the ballot theorem we are given an expression where the total number of ballot permutations $\binom{a+b}{a}$ is multiplied by the fraction $(a - kb)/(a + b)$. Dvoretzky and Motzkin [12] solve the ballot problem by introducing the *cycle lemma* which makes evident the reason for the fraction. The cycle lemma provides a surprising result: for any ballot sequence of a votes for A and b votes for B , exactly $a - kb$ of the $a + b$ cyclic permutations of the sequence are good. Consequently, a fraction of $(a - kb)/(a + b)$ of all ballot permutations

are good.

Dershowitz and Zaks [11] give two elegant proofs of the cycle lemma. Their first proof is a generalization (and simplification) of the proofs in [7], [25], and [26]; their second proof follows [15], [22], and [32]. In the following proof of the ballot theorem, we include what is essentially their first proof of the cycle lemma.

Proof 3. We can express a ballot permutation as a sequence of $a + b$ terms where each term is either 1 or $-k$; votes for A correspond to the 1's and votes for B correspond to the $-k$'s. A sequence is called *good* if every partial sum is positive, and *bad* otherwise. Observe that the sum of a sequence is $a - kb \geq 0$.

Let C be any circular arrangement of a 1's and b $-k$'s. We now prove the cycle lemma: of the $a + b$ terms in C , exactly $a - kb$ start good sequences when C is read once around clockwise.

By the pigeonhole principle there must exist a sequence $X = 1, 1, \dots, 1, -k$ in C with k consecutive 1's. No term of X can start a good sequence, for when we get to the $-k$ we would have a partial sum less than or equal to zero.

Let C' be the circular arrangement created from C by removing X . Since the sequence X has sum 0, it has no "net effect" on good sequences. Thus, a term of C starts a good sequence if and only if the corresponding term in C' starts a good sequence. Consequently, C and C' have exactly the same number of terms that start good sequences. Continuing in this manner, one removes sequences of the form $1, 1, 1, \dots, 1, -k$ until a circular arrangement consisting only of 1's remains. At this stage, there are $a - kb$ 1's, and every

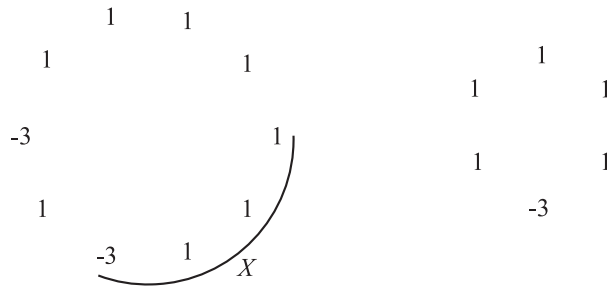


Figure 2: Example of C and C' with $k = 3$.

term starts a good sequence. Hence there are exactly $a - kb$ good sequences in C , and the cycle lemma is proved.

If there is periodicity in C then not all $a - kb$ good sequences will be distinct. However, we can conclude that the *ratio* of good sequences to all sequences is $a - kb$ to $a + b$. Therefore, the number of good sequences is

$$\frac{a - kb}{a + b} \binom{a + b}{a}. \quad \square$$

Dvoretzky and Motzkin [12] state and prove the cycle lemma as a means of solving the ballot problem, but Dershowitz and Zaks [11] point out that this is a “frequently rediscovered combinatorial lemma” and they provide two other applications of the lemma. They write,

The Cycle Lemma is the combinatorial analogue of the Lagrange inversion formula; see Raney [22], Cori [10] and Gessel [13]. Other proofs of varying degree of generality may be found in Dvoretzky and Motzkin [12] (discussed in Grossman [15]), Motzkin [21] (two proofs), Hall [16], Raney [22], Yaglom and Yaglom [32], Takács [29], Silberger [25], Bergman [7] (three proofs), Sands [24] and

Singmaster [26]. (The first paper [12] is not credited by the other authors, but is referenced in Barton and Mallows [6] and Mohanty [20].) Dvoretzky and Motzkin, Motzkin, and Yaglom and Yaglom give the lemma in its general form; the other papers prove only the case $k = 1$ or $a - kb = 1$. Generalizations of the Cycle Lemma to non-integer k and sequences of reals may be found in Dvoretzky and Motzkin [12] and Spitzer [27], respectively.

(Reference numbers and notation in the above quote have been modified for consistency with this paper.)

Proof 4: A Uniform Partition

Consider the $\binom{2n}{n}$ possible lattice paths starting from the origin and consisting of n upsteps $(1, 1)$ and n downsteps $(1, -1)$. It turns out, surprisingly, that the number of these paths with i upsteps above the x -axis ($0 \leq i \leq n$) is the same, regardless of the value of i . Consequently, the number of paths with all n upsteps above the x -axis must be $\binom{2n}{n}/(n+1)$. This fact is often called the Chung-Feller theorem [9, Thm. 2A]; however, it was actually given 1909 by MacMahon [18, p. 167, §20] in the process of solving the ballot problem (for $k = 1$) via the theory of partitions.

In the following proof we apply a similar approach by creating a set Ψ with $(a - kb)\binom{a+b}{a}$ elements, and partitioning this set into $a + b$ subsets of equal size (that is, we uniformly partition Ψ into $a + b$ subsets). One of the subsets corresponds to the set of good ballot permutations, and from this we can conclude the ballot theorem. It appears that the following proof is the

first to prove the ballot theorem by means of a uniform partition. It is based on and extends the proofs found in [31].

Proof 4. Consider lattice paths starting from the origin and consisting of a upsteps $(1, 1)$ and b downsteps $(1, -k)$, and assume the strict inequality $a > kb$. Let $\mathcal{A} = \mathcal{A}(a, b, k)$ be the set of all such paths. Given path $P \in \mathcal{A}$ we let $L(P)$ denote the set of x -values of the $a - kb$ “rightmost lowest” vertices of P ; see Figure 3. More precisely, given path $P \in \mathcal{A}$, let y_0 denote the least y -value of all the vertices of P , and let $r(t)$ denote the x -value of the rightmost vertex of P along the line $y = t$; then $L(P) = \{r(t) \mid t \in \mathbb{Z}, y_0 \leq t \leq y_0 + (a - kb) - 1\}$.

Let $\Psi = \{(P, j) \mid P \in \mathcal{A}, j \in L(P)\}$ and note that $|\Psi| = (a - kb) |\mathcal{A}|$. Let $\Omega_i = \{(P, i) \in \Psi \mid i \in L(P)\}$, defined for $0 \leq i \leq a + b - 1$. The sets Ω_i partition Ψ into $a + b$ disjoint subsets.

Claim 1: There is a one-to-one correspondence between Ω_0 and the set of good paths. If $P \in \mathcal{A}$ is good, then $(0, 0)$ is the lowest vertex in P and it is the only vertex on the x -axis, so $(P, 0) \in \Omega_0$. Conversely, if $(P, 0)$ is in Ω_0 , then no vertex of P can lie on the x -axis to the right of the origin, and so P is good.

Claim 2: The sets Ω_i uniformly partition Ψ . We show this by providing a one-to-one correspondence between Ω_i and Ω_0 . If $(P, i) \in \Omega_i$, then write $P = XY$ where X is the initial path of P consisting of the first i steps, and Y consists of the remaining steps. Since $i \in L(P)$, we can observe that Y stays above the height of its initial vertex, and X never descends $a - kb$ or more units below the height of its terminal vertex. Consequently the path YX is good and $(YX, 0) \in \Omega_0$. Conversely, if $(Q, 0) \in \Omega_0$, then write $Q = YX$

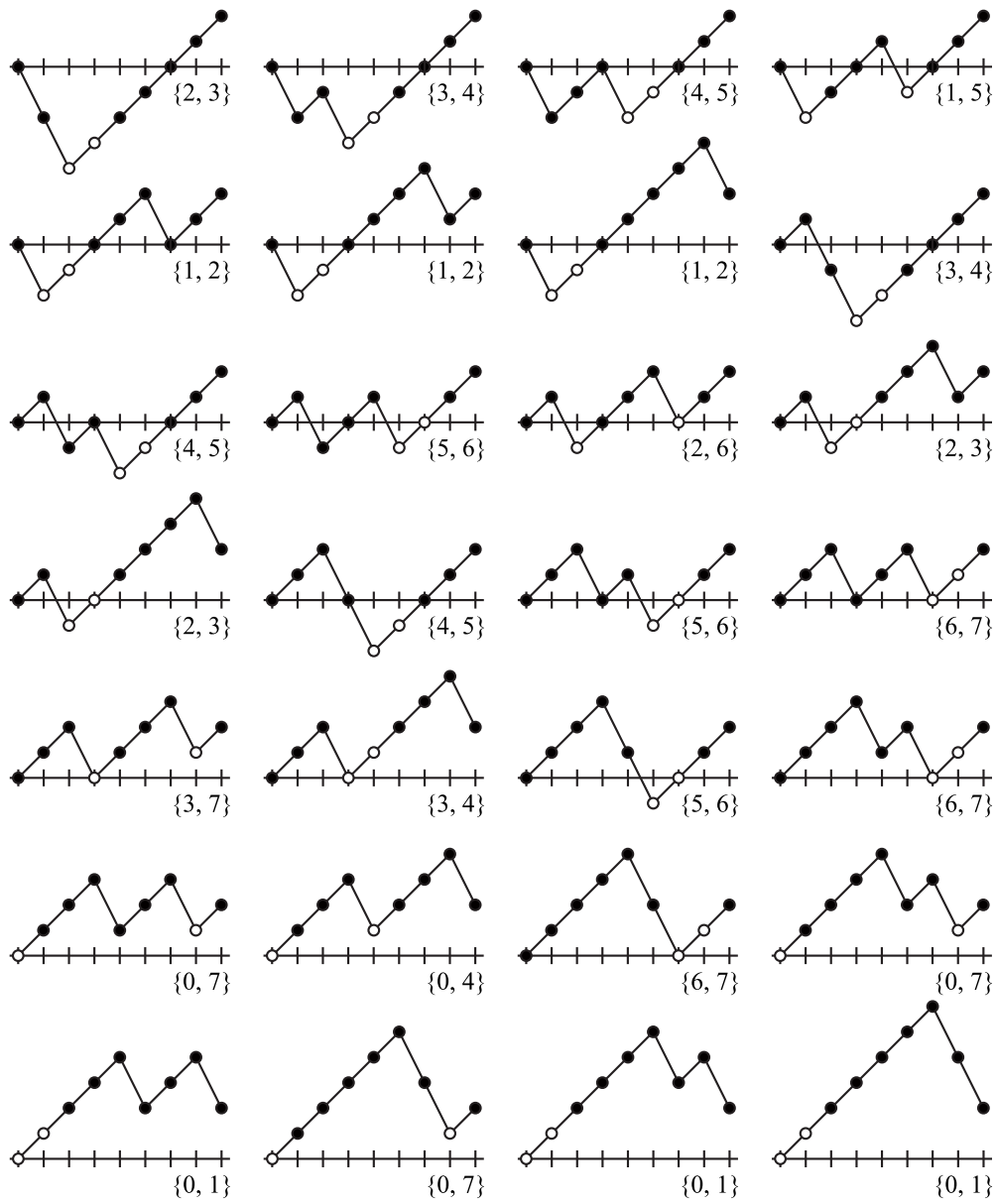


Figure 3: Example with $k = 2$, $a = 6$, $b = 2$. For each path $P \in \mathcal{A}$, the path P and the set $L(P)$ are shown. Observe that among all the sets $L(P)$, each number from 0 to 7 occurs exactly 7 times.

where X consists of the final i steps of Q . The same qualities of X and Y hold as noted above, and the pair $(XY, i) \in \Omega_i$.

The two claims above imply that the number of good paths in \mathcal{A} is

$$|\Omega_0| = \frac{|\Psi|}{a+b} = \frac{(a-kb)|\mathcal{A}|}{a+b} = \frac{a-kb}{a+b} \binom{a+b}{a}. \quad \square$$

Suppose we let \mathcal{A}_i denote the set of paths for which i is among the x -values of the $a-kb$ rightmost vertices, i.e., $\mathcal{A}_i = \{P \in \mathcal{A} \mid i \in L(P)\}$. When $a-kb = 1$, the sets $\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_{a+b-1}$ are disjoint and all have the same cardinality. In other words, partitioning \mathcal{A} according to the x -value of a path's rightmost lowest vertex creates a uniform partition of \mathcal{A} .

Curiously, when we allow $a-kb \geq 1$, the sets \mathcal{A}_i continue to have the same cardinality. However, they are no longer disjoint. To the contrary, each path in \mathcal{A} will be a member of precisely $a-kb$ of these sets.

The Weak Ballot Problem, Catalan Numbers

The ballot problem is often stated in a “weak” version: suppose that candidate A receives m votes and candidate B receives n votes, where $m \geq kn$ for some positive integer k , and compute the number of ways the ballots can be ordered so that A always has *at least* k times as many votes as B throughout the counting of the ballots.

Any ballot permutation in which A maintains *at least* k times the number of votes for B can be converted into one in which A has *more than* k times the number of votes for B by simply appending a vote for A to the beginning of the permutation. Clearly this process is reversible, and hence the solution to the weak version is the same as the “strict” version when A receives $m+1$

votes and B receives n votes:

$$\frac{(m+1) - kn}{(m+1) + n} \binom{(m+1) + n}{m+1} = \frac{m+1 - kn}{m+1} \binom{m+n}{m}.$$

Putting $k = 1$ and $m = n$ produces the well-known *Catalan numbers*

$$C_n = \frac{1}{n+1} \binom{2n}{n}.$$

Requiring only that $m = kn$ produces the *generalized Catalan numbers*, also called the *k-Catalan numbers*

$$C_n^k = \frac{1}{kn+1} \binom{(k+1)n}{n}.$$

The interested reader should see [28, pp. 219–229] and Stanley’s website <http://www-math.mit.edu/~rstan/ec/> for an extensive list of combinatorial interpretations of the Catalan numbers. Furthermore, see [17] for several interpretations of the generalized Catalan numbers.

Acknowledgments

I would like to extend sincere thanks to the two referees who offered many valuable suggestions.

References

- [1] J. Aebly, Démonstration du problème du scrutin par des considérations géométriques, *L’enseignement mathématique* **23** (1923) 185–186.
- [2] A. Aeppli, A propos de l’interprétation géométrique du problème du scrutin, *L’enseignement mathématique* **23** (1923) 328–329.

- [3] A. Aeppli, *Zur Theorie verketteter Wahrscheinlichkeiten, Markoff'sche Ketten höherer Ordnung*, Ph.D. Thesis, Eidgenössische Technische Hochschule, Zürich, 1924.
- [4] D. André, Solution directe du problème résolu par M. Bertrand, *Comptes Rendus de l'Académie des Sciences*, Paris **105** (1887) 436–437.
- [5] É. Barbier, Généralisation du problème résolu par M. J. Bertrand, *Comptes Rendus de l'Académie des Sciences*, Paris **105** (1887) p. 407.
- [6] D.E. Barton and C.L. Mallows, Some aspects of the random sequence, *Ann. Math. Stat.* **36** (1965) 236–260.
- [7] G.M. Bergman, Terms and cyclic permutations, *Algebra Universalis* **8** (1978) 129–130.
- [8] J. Bertrand, Solution d'un problème, *Comptes Rendus de l'Académie des Sciences*, Paris **105** (1887) p. 369.
- [9] K.L. Chung and W. Feller, On fluctuations in coin tossing, *Proc. Natl. Acad. Sci. USA* **35** (1949) 605–608.
- [10] R. Cori, in: *Combinatorics on Words* (M. Lothaire, ed.), Encyclopedia of Mathematics and Its Applications, vol. 17, Addison-Wesley, Reading, Massachusetts, 1983.
- [11] N. Dershowitz and S. Zaks, The cycle lemma and some applications, *Europ. J. Combinatorics* **11** (1990) 35–40.
- [12] A. Dvoretzky and Th. Motzkin, A problem of arrangements, *Duke Math. J.* **14** (1947) 305–313.

- [13] I.M. Gessel, A combinatorial proof of the multivariate Lagrange inversion formula, *J. Combin. Theory Ser. A* **45** (1987) 178–195.
- [14] I.P. Goulden and L.G. Serrano, Maintaining the spirit of the reflection principle when the boundary line has arbitrary integer slope, *J. Comb. Theory, Series A* **104** (2003) 317–326.
- [15] H.D. Grossman, Fun with lattice points—21, *Scripta math.* **16** (1950) 120–124.
- [16] P. Hall, Some word problems, *J. London Math. Soc.* **33** (1958) 482–496.
- [17] P. Hilton and J. Pedersen, Catalan numbers, their generalizations, and their uses, *Math. Intell.* **13** (1991) 64–75.
- [18] P.A. MacMahon, Memoir on the theory of the partitions of numbers. part iv: on the probability that the successful candidate at an election by ballot may never at any time have fewer votes than the one who is unsuccessful; on a generalization of this question; and on its connexion with other questions of partition, permutation, and combination, *Philosophical Transactions of the Royal Society of London, Series A* **209** (1909) 153–175. Also *Collected Papers Vol. 1*, (G.E. Andrews, ed.), MIT Press, Cambridge, Mass 1978, 1292–1314.
- [19] D. Mirimanoff, A propos de l'interprétation géométrique du problème du scrutin, *L'enseignement mathématique* **23** (1923) 187–189.
- [20] S.G. Mohanty, *Lattice Path Counting and Applications*, Academic Press, New York, 1979.

- [21] Th. Motzkin, Relations between hypersurface cross ratios and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products, *Bull. Am. Math. Soc.* **54** (1948) 352–360.
- [22] G.M. Raney, Functional composition patterns and power series reversion, *Trans. Am. Math. Soc.* **94** (1960) 441–451.
- [23] M.S. Renault, Lost (and found) in translation: André’s actual method and its application to the generalized ballot problem, *Amer. Math. Monthly* to appear. See webpace.ship.edu/msrenault/ballotproblem/.
- [24] A.D. Sands, On generalized Catalan numbers, *Discr. Math.* **21**(2) (1978) 219–221.
- [25] D.M. Silberger, Occurrences of the integer $(2n - 2)!/n!(n - 1)!$, *Roczniki Polskiego Towarzystwa Math. I* **13** (1969), 91–96.
- [26] D. Singmaster, An elementary evaluation of the Catalan numbers, *Am. Math. Monthly* **85** (1978) 366–368.
- [27] F. Spitzer, A combinatorial lemma and its application to probability theory, *Trans. Am. Math. Soc.* **82** (1956) 323–339.
- [28] R.P. Stanley, *Enumerative Combinatorics, Vol. 2*, Cambridge University Press, Cambridge, UK, 1999.
- [29] L. Takács, *Combinatorial Methods in the Theory of Stochastic Processes*, John Wiley, New York, 1967.

- [30] L. Takács, On the ballot theorems, *Advances in Combinatorial Methods and Applications to Probability and Statistics*, Birkhäuser, 1997.
- [31] W. Woan, Uniform partitions of lattice paths and Chung-Feller generalizations, *Amer. Math. Monthly* **108** (2001) 556–559.
- [32] A.M Yaglom and I.M. Yaglom, *Challenging Mathematical Problems with Elementary Solutions, vol 1: Combinatorial Analysis and Probability Theory*, Holden Day, San Francisco, 1964.