Catalan Numbers, Their Generalization, and Their Uses

Peter Hilton and Jean Pedersen

Probably the most prominent among the special integers that arise in combinatorial contexts are the binomial coefficients $\binom{n}{r}$. These have many uses and, often, fascinating interpretations [9]. We would like to stress one particular interpretation in terms of paths on the integral lattice in the coordinate plane, and discuss the celebrated *ballot problem* using this interpretation.

A path is a sequence of points $P_0 P_1 \ldots P_m$, $m \geqslant 0$, where each P_i is a lattice point (that is, a point with *integer* coordinates) and P_{i+1}, $i \geqslant 0$, is obtained by stepping one unit east or one unit north of P_i. We say that this is a *path from P to Q* if $P_0 = P$, $P_m = Q$. It is now easy to count the number of paths.

THEOREM 0.1. *The number of paths from* (0,0) *to* (a,b) *is the binomial coefficient* $\binom{a+b}{a}$.

Proof: We may denote a path from (0,0) to (a,b) as a sequence consisting of a Es (E for East) and b Ns (N for North) in some order. Thus the number of such paths

Peter Hilton

Jean Pedersen

Peter Hilton is Distinguished Professor of Mathematics at the State University of New York at Binghamton. He is the author of numerous books and research articles, mainly in the areas of algebraic topology, homological algebra, and group theory. He has a long-standing interest in mathematics education—he was chairman of the United States Commission on Mathematical Instruction and chairman of the National Science Council Committee on Applied Mathematics Training—and has been much concerned, with Jean Pedersen, to revive the study of geometry at the pre-college and college levels and to develop an integrated mathematics curriculum.

Jean Pedersen has been a member of the faculty of the Mathematics Department of Santa Clara University, California, for twenty-four years. Her principal research interests are in polyhedral geometry and combinatorics, where her work has been recognized by invitations to give addresses at international conferences (e.g., The Coxeter Symposium in Toronto, The Escher Congress in Rome, The Fifth International Conference on Geometry in Haifa). Peter Hilton and she have collaborated extensively on mathematical research and in the field of mathematics education; they have published jointly over forty articles in the past ten years and two books, *Fear No More* and *Build Your Own Polyhedra*.

is the number of such sequences. A sequence consists of $(a + b)$ symbols and is determined when we have decided which a of the $(a + b)$ symbols should be Es. Thus we have $\binom{a+b}{a}$ choices of symbol.

The Ballot Problem

In the late nineteenth century the so-called *ballot problem* was exercising the minds of mathematicians and probabilists. We suppose an election is held in which there are two candidates, A and B. We further suppose that A receives a votes, B receives b votes with $a > b$, so that A is elected. The question raised is this—what is the probability that, throughout the counting of votes, A stays ahead of B? This translates easily to a question about paths on the integral lattice in the coordinate plane. Each path from (0,0) to (a,b) represents a possible sequence of counting of votes, so there are

$$\binom{a+b}{a}$$

such paths, and we need to know how many of these paths remain below the line $y = x$ except at the initial point (0,0). This number is the number of 2-*good paths* from (1,0) to (a,b), as defined later.

We will present another family of special integers, called the *Catalan numbers* (after the nineteenth-century Belgian[1] mathematician Eugène Charles Catalan; see the historical note at the end of this article), which are closely related, both algebraically and conceptually, to the binomial coefficients. These Catalan numbers have many combinatorial interpretations, of which we will emphasize three, in addition to their interpretation in terms of 2-good paths on the integral lattice. Because all these interpretations remain valid if one makes a conceptually obvious generalization of the original definition of Catalan numbers, we will present these interpretations in this generalized form. Thus the definitions we are about to give depend on a parameter p, which is an integer greater than one. The original Catalan numbers—or, rather, characterizations of these numbers—are obtained by taking $p = 2$.

Let us then present three very natural combinatorial concepts, which arise in rather different contexts, but which turn out to be equivalent.

Let p be a fixed integer greater than 1. We define three sequences of positive integers, depending on p, as follows:

[1] He lived his life in Belgium; but, being born in 1814, he was not born a Belgian!

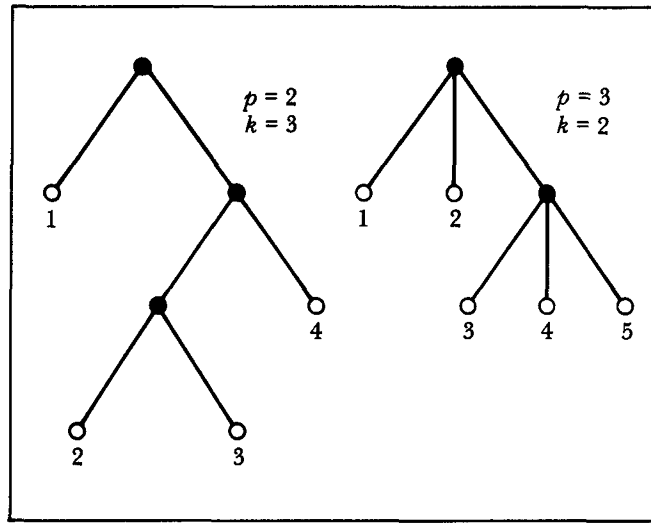

Figure 1. Left: a binary (or 2-ary) tree with 3 source-nodes (●) and 4 end-nodes (○) and, at right, a ternary (or 3-ary) tree with 2 source-nodes (●) and 5 end-nodes (○).

$p = 2$, $k = 3$	$p = 3$, $k = 2$
$(s_1((s_2\ s_3)s_4))$	$(s_1\ s_2(s_3\ s_4\ s_5))$

Figure 2. Left: an expression involving 3 applications of a binary operation applied to 4 symbols and, at right, an expression involving 2 applications of a ternary operation applied to 5 symbols.

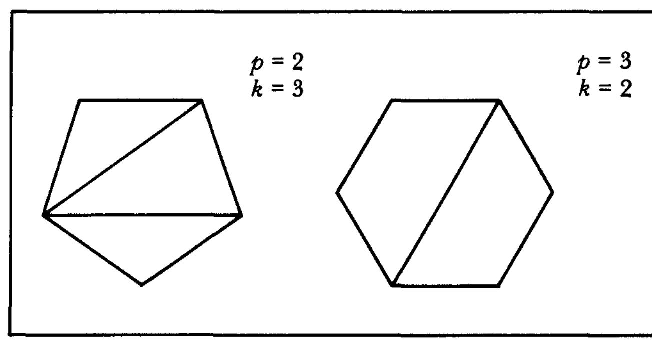

Figure 3. Left: a 5-gon subdivided into three 3-gons by 2 diagonals and, at right, a 6-gon subdivided into two 4-gons by 1 diagonal.

a_k: ${}_pa_0 = 1$, ${}_pa_k$ = the number of p-ary trees with k source-nodes, $k \geq 1$ (see Figure 1).

b_k: ${}_pb_0 = 1$, ${}_pb_k$ = the number of ways of associating k applications of a given p-ary operation, $k \geq 1$ (see Figure 2).

c_k: ${}_pc_0 = 1$, ${}_pc_k$ = the number of ways of subdividing a convex polygon into k disjoint $(p + 1)$-gons by means of non-intersecting diagonals, $k \geq 1$ (see Figure 3).

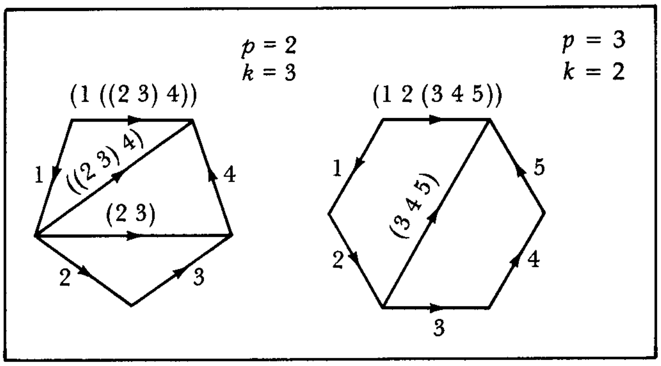

Figure 4. The dissected polygons of Figure 3, with diagonals and last side labeled.

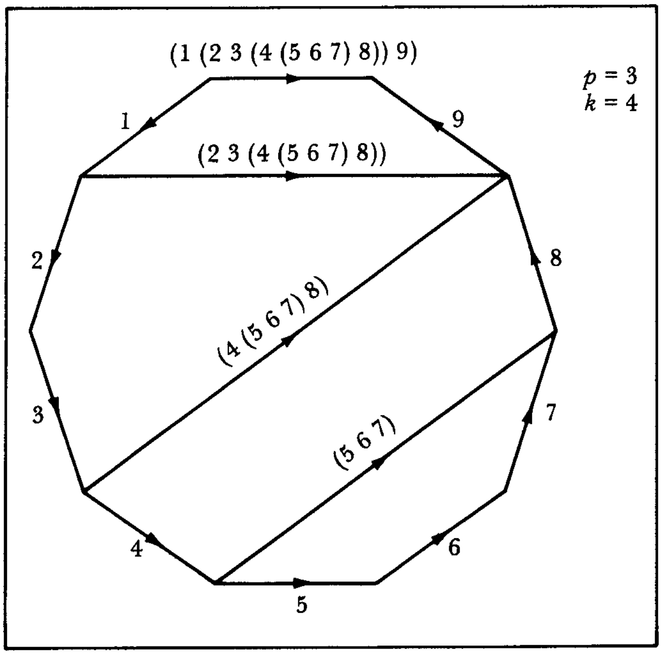

Figure 5. The polygonal dissection associated with $(s_1(s_2s_3(s_4(s_5s_6s_7)s_8))s_9)$.

(As indicated, we suppress the 'p' from the symbol if no ambiguity need be feared.)

Note that, if $k \geqslant 1$,

(i) a p-ary tree with k source-nodes has $(p - 1)k + 1$ end-nodes and $pk + 1$ nodes in all (Figure 1);

(ii) the k applications of a given p-ary operation are applied to a sequence of $(p - 1)k + 1$ symbols (Figure 2);

(iii) the polygon has $(p - 1)k + 2$ sides and is subdivided into k disjoint $(p + 1)$-gons by $(k - 1)$ diagonals (Figure 3).

A well-known and easily proved result (for the case $p = 2$ see Sloane [8]) is the following:

THEOREM 0.2. ${}_pa_k = {}_pb_k = {}_pc_k$.

The reader will probably have no trouble seeing how the trees of Figure 1 are converted into the corresponding expressions of Figure 2—and vice versa. However, relating the trees of Figure 1 (or the parenthetical expressions of Figure 2) to the corresponding dissected polygons of Figure 3 is more subtle (and the hint given in [8], in the case $p = 2$, is somewhat cryptic), so we will indicate the proof that $b_k = c_k$. Thus, suppose that we are given a rule for associating k applications of a given p-ary operation to a string of $(p - 1)k + 1$ symbols $s_1, s_2, \ldots, s_{(p-1)k+1}$. Label the successive sides (in the anticlockwise direction) of the convex $((p - 1)k + 2)$-gon with the integers from 1 to $(p - 1)k + 1$, leaving the top, horizontal side unlabeled. Pick the first place along the expression (counting from the left) where a succession of p symbols is enclosed in parentheses. If the symbols enclosed run from s_{j+1} to s_{j+p}, draw a diagonal from the initial vertex of the $(j + 1)^{\text{st}}$ side to the final vertex of the $(j + p)^{\text{th}}$ side, and label it $(j + 1, \ldots, j + p)$ Also imagine the part $(s_{j+1} \cdots s_{j+p})$ replaced by a single symbol. We now have effectively reduced our word to a set of $(k - 1)$ applications and our polygon to a set of 2 polygons, one a $(p + 1)$-gon, the other a $[(p - 1)(k - 1) + 2]$-gon, so we proceed inductively to complete the rule for introducing diagonals. Moreover, the eventual label for the last side will correspond precisely to the string of $(p - 1)k + 1$ symbols with the given rule for associating them, that is, to the original expression. Figure 4 illustrates the labeling of the dissections of the pentagon and hexagon that are determined by the corresponding expressions of Figure 2. The converse procedure, involving the same initial labeling of the original sides of a dissected polygon, and the successive labeling of the diagonals introduced, leads to an expression that acts as label for the last (horizontal) side.

We now illustrate a slightly more complicated situation:

Example 0.1. Let $p = 3$, $k = 4$ and consider the expression

$$s_1(s_2s_3(s_4(s_5s_6s_7)s_8))s_9).$$

The dissection of the convex 10-gon associated with this expression, with the appropriate labeling, is shown in Figure 5. The corresponding ternary tree is shown in Figure 6, which also demonstrates how the tree may be associated directly with the dissected polygon. The rule is "Having entered a $(p + 1)$-gon through one of its sides, we may exit by any of the other p sides."

In the case $p = 2$, any of a_k, b_k, c_k may be taken as the definition of the k^{th} Catalan number. Thus we may regard any of ${}_pa_k$, ${}_pb_k$, ${}_pc_k$ as defining the *generalized* k^{th} *Catalan number*. Moreover, the following result is known (see [6]).

THEOREM 0.3. ${}_pa_k = \dfrac{1}{k}\dbinom{pk}{k-1} = \dfrac{1}{(p-1)k+1}\dbinom{pk}{k}$, $k \geqslant 1$.

The "classical" proof of this result is rather sophisticated. One of our principal aims is to give an elementary proof of Theorem 0.3, based on yet a fourth interpretation of the generalized Catalan numbers which, like that of the binomial coefficients given earlier, is also in terms of paths on the integral lattice. We lay

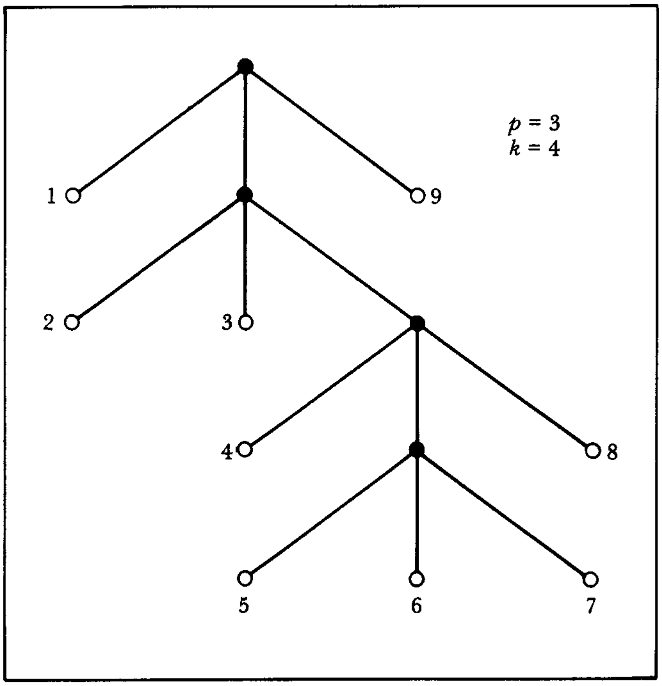

Figure 6 (a). **The tree associated with $(s_1(s_2s_3(s_4(s_5s_6s_7)s_8))s_9)$.**

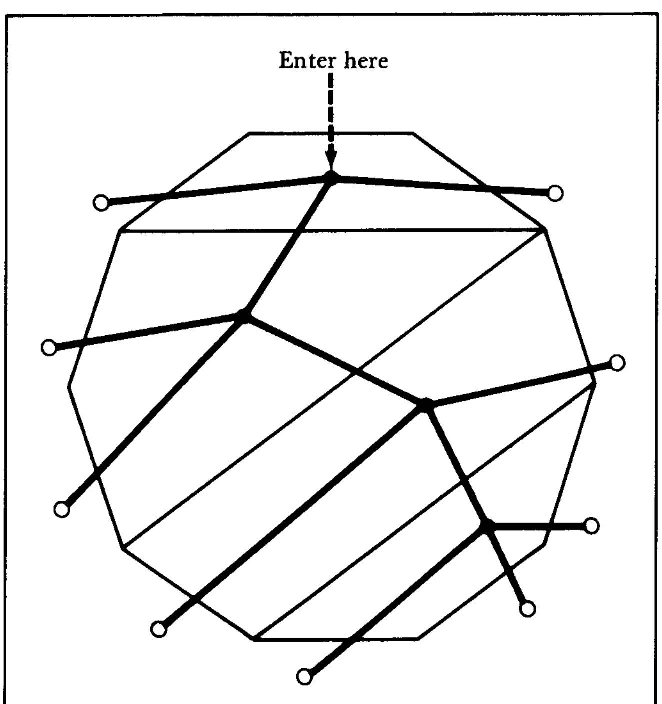

Figure 6 (b). Associating a tree with a dissected polygon (in this case, the tree shown in (a) and the polygon of Figure 5). Note that • = interior node = source-node; ○ = exterior node = end node.

particular stress on the flexibility of this fourth interpretation, which we now describe.

We say that a path from P to Q is *p-good* if it lies entirely below the line $y = (p - 1)x$. Let $d_k = {}_pd_k$ be the number of p-good paths from $(0, -1)$ to $(k, (p - 1)k - 1)$. (Thus, by our convention, $d_0 = 1$.) We extend Theorem 0.2 to assert the following.

THEOREM 0.4. ${}_pa_k = {}_pb_k = {}_pc_k = {}_pd_k$.

Proof: First we restate (with some precision) the definitions of b_k and d_k, which numbers we will prove equal.

b_k = the number of expressions involving k applications of a p-ary operation (on a word of $k(p - 1) + 1$ symbols). Let $\underline{E}$ be the set of all such expressions.

d_k = the number of good paths from $(0, -1)$ to $(k, (p - 1)k - 1)$; by following such a path by a single vertical segment, we may identify such paths with certain paths (which we will also call 'good') from $(0, -1)$ to $(k, (p - 1)k)$. Let $\underline{P}$ be the set of all such paths.

We set up functions Φ: $\underline{E} \to \underline{P}$, Ψ: $\underline{P} \to \underline{E}$, which are obviously mutual inverses. Thus Φ is obtained by removing all final parentheses[2] and (reading the resulting expression from left to right) interpreting a parenthesis as a horizontal move and a symbol as a vertical move. The inverse rule is Ψ. It remains to show (i) that ImΦ consists of elements of $\underline{P}$ (and not merely of arbitrary paths from $(0, -1)$ to $(k, (p - 1)k)$; and (ii) that ImΨ consists of elements of $\underline{E}$ (that is, of well-formed, meaningful expressions).

(i) Let E be an expression. We argue by induction on k that ΦE is a good path. Plainly ΦE is a path from $(0, -1)$ to $(k, (p - 1)k)$, so we have only to prove it good. This obviously holds if $k = 1$. Let $k \geq 2$ and let E involve k applications. There must occur somewhere in E the section

$$(s_{j+1}s_{j+2} \cdots s_{j+p})$$

Call this the *key* section. We wish to show that if (u,v) is a point on the path ΦE that is *not* the endpoint, then $v < (p - 1)u$. Let E' be the expression obtained from E by replacing the key section by s. Then E' involves $(k - 1)$ applications. Now if A is the point on ΦE corresponding to the parenthesis of the key section, then the inductive hypothesis tells us that $v < (p - 1)u$ if (u,v) occurs prior to A. Assume now that (u,v) is not prior to A. There are then 2 possibilities:

[2] This converts our expression into a meaningful reverse Polish expression (MRPE); see [2]. Note that the closing parentheses are, strictly speaking, superfluous, provided that we know that we are dealing with a p-ary operation. It is not even necessary to know this if we are sure that the expression is well-formed, since the 'arity' is $\frac{\sigma-1}{\kappa} + 1$, where σ is the number of symbols and κ is the number of opening parentheses.

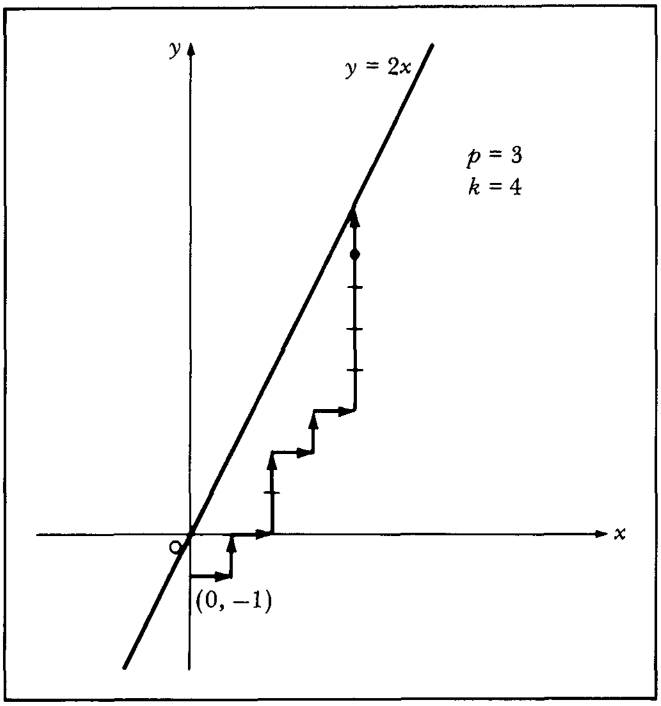

Figure 7 (a). The path associated with the expression $E = (s_1(s_2s_3(s_4(s_5s_6s_7)s_8))s_9)$.

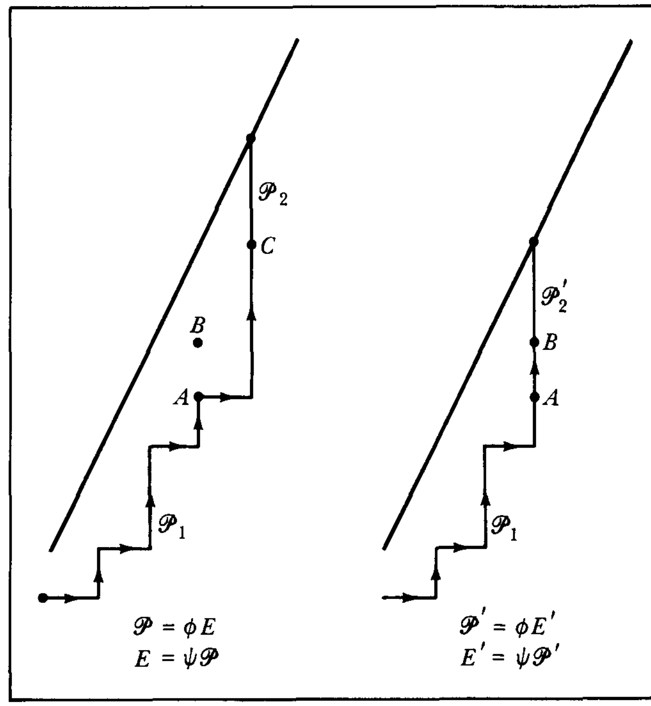

Figure 7 (b). The inductive step, in either direction.

If E ends with s_{j+p}, so that E' ends with s, then $\exists(u',v')$, the last point of $\Phi E'$, such that $u' = u - 1, v' > v - p + 1, v' = (p - 1)u'$.

Hence $v - p + 1 < (p - 1)(u - 1)$, so $v < (p - 1)u$.

If E does not end with s_{j+p}, so that E' does not end with s, then $\exists(u',v')$, not the last point of $\Phi E'$, such that $u' = u - 1$, $v' \geqslant v - p + 1$, $v' < (p - 1)u'$ (by the inductive hypothesis). Hence $v - p + 1 < (p - 1)(u - 1)$, so $v < (p - 1)u$.

This completes the induction in the Φ-direction. Figure 7, which gives a special, but not particular, case, may be helpful in following the argument.

(ii) Let $\mathcal{P}$ be a good path to $(k, (p - 1)k)$. We argue by induction on k that $\Psi\mathcal{P}$ is a well-formed expression. This holds if $k = 1$, because then there is only one good path $\mathcal{P}$ and $\Psi\mathcal{P}$ is $(s_1s_2 \ldots s_p)$. Let $k \geqslant 2$. Then because the path climbs altogether $k(p - 1) + 1$ places in k jumps, there must be a jump somewhere (not at the initial point) of at least p places. Let A be a point on $\mathcal{P}$ where the path takes one horizontal step followed by p vertical steps, bringing it to C. Let B be the point one step above A (so that B is not on $\mathcal{P}$). Then B is on $y = (p - 1)x$ if and only if C is on $y = (p - 1)x$, that is, if and only if C is the endpoint of $\mathcal{P}$; otherwise B is below $y = (p - 1)x$. Let $\mathcal{P}_1$, $\mathcal{P}_2$ be the parts of $\mathcal{P}$ ending in A and beginning in C, respectively. Let $\mathcal{P}_2'$ be the translate of $\mathcal{P}_2$, given by

$$u' = u - 1, v' = v - (p - 1);$$

and let $\mathcal{P}'$ be the path consisting of $\mathcal{P}_1$, followed by AB, followed by $\mathcal{P}_2'$. It is trivial that $\mathcal{P}'$ is a good path to $(k - 1, (p - 1)(k - 1))$. Thus, by the inductive hypothesis, $\Psi\mathcal{P}'$ is a well-formed expression involving $(k - 1)$ applications. Let s be the symbol in $\Psi\mathcal{P}'$ corresponding to AB. Replace s by $(s_{j+1}s_{j+2} \ldots s_{j+p}$, where these symbols do not occur in $\Psi\mathcal{P}'$. Then the new expression is well-formed and is obviously $\Psi\mathcal{P}$. This completes the induction in the Ψ-direction so that Theorem 0.4 is established.

With Theorem 0.4 established, we may proceed to calculate d_k. This is achieved in Section 2; a more general algorithm for counting p-good paths is described in Section 3.

By considering the last application of the given p-ary operation, it is clear that ${}_pb_k$ satisfies the recurrence relation (generalizing the familiar formula in the case $p = 2$)

$$ {}_pb_k = \sum_{i_1+i_2+\,\cdot\,+i_p=k-1} {}_pb_{i_1}\,{}_pb_{i_2} \cdots {}_pb_{i_p}, \; k \geqslant 1. \quad (0.1)$$

In Section 2 we also enunciate a similar recurrence relation for ${}_pd_k$, thus providing an alternative proof that ${}_pb_k = {}_pd_k$. We will also show in Section 2, in outline, how (0.1) leads to a proof of Theorem 0.3 via generating functions and the theory of Bürmann-Lagrange series (see [5]).

The case $p = 2$, dealing with the Catalan numbers as originally defined, is discussed in Section 1. In that special case we have available an elegant proof that ${}_2d_k = \frac{1}{k}\binom{2k}{k-1}$, $k \geqslant 1$, using André's Reflection Method [1]; however, we have not discovered in the literature a means of generalizing this method effectively.

The authors wish to thank Richard Guy, Richard Johnsonbaugh, Hudson Kronk, and Tom Zaslavsky for very helpful conversations and communications during the preparation of this article.

1. Catalan Numbers and the Ballot Problem

The ballot problem (see box) was first solved by Bertrand, but a very elegant solution was given in 1887 by the French mathematician Désiré André (see [1], [4]). In fact, his method allows one easily to count the number of 2-good paths from (c,d) to (a,b), where (c,d), (a,b) are any two lattice points below the line $y = x$. Because $p = 2$ throughout this discussion, we will speak simply of *good* paths; any path from (c,d) to (a,b) in the complementary set will be called a *bad* path. We will assume from the outset that both good paths and bad paths exist; it is easy to see that this is equivalent to assuming

$$d < c \leqslant b < a.$$

(Note that this condition is certainly satisfied in our ballot problem, provided the losing candidate receives at least 1 vote!)

Now a simple translation argument shows that the total number of paths from $P(c,d)$ to $Q(a,b)$ is the binomial coefficient

$$\binom{(a+b)-(c+d)}{a-c};$$

thus it suffices to count the number of *bad* paths from P to Q; see Figure 8. If $\mathcal{P}$ is a bad path, let it first make contact with the line $y = x$ at F; let $\mathcal{P}_1$, $\mathcal{P}_2$ denote the subpaths PF, FQ, so that, using juxtaposition for path-composition, $\mathcal{P} = \mathcal{P}_1\mathcal{P}_2$; and let $\overline{\mathcal{P}}_1$ be the path obtained from $\mathcal{P}_1$ by *reflecting in the line* $y = x$. Then, if $\overline{\mathcal{P}} = \overline{\mathcal{P}}_1\mathcal{P}_2$, $\overline{\mathcal{P}}$ is a path from $\overline{P}(d,c)$ to $Q(a,b)$; and the rule $\mathcal{P} \rightarrow \overline{\mathcal{P}}$ sets up a one-one correspondence between the set of bad paths from P to Q and the set of paths from $\overline{P}$ to Q. It follows that there are $\binom{(a+b)-(c+d)}{a-d}$ bad paths from P to Q, and hence

$$\binom{(a+b)-(c+d)}{a-c} - \binom{(a+b)-(c+d)}{a-d} \qquad (1.1)$$

good paths from P to Q. The argument, illustrated in Figure 8, is called *André's Reflection Method.*

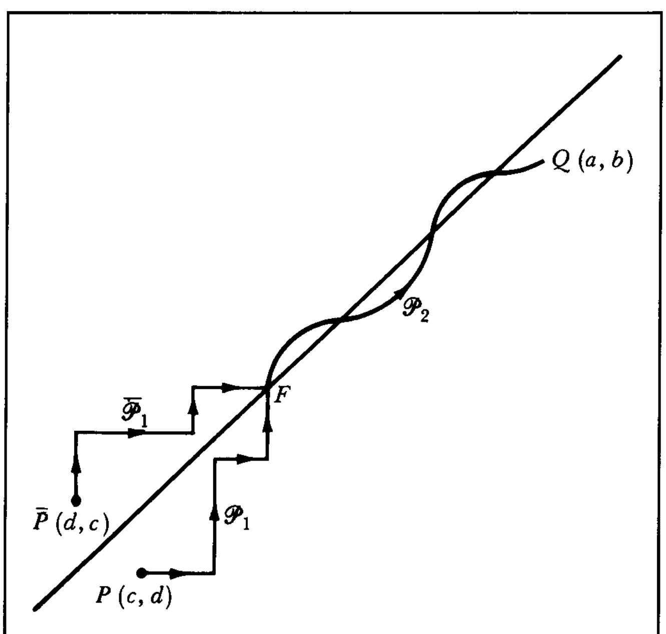

Figure 8. André's reflection method.

Notice, in particular, that if $c = 1$, $d = 0$, this gives the number of good paths from (1,0) to (a,b) as

$$\binom{a+b-1}{a-1} - \binom{a+b-1}{a} = \frac{a-b}{a+b}\binom{a+b}{a}.$$

Thus the probability that a path from (0,0) to (a,b) proceeds first to (1,0) and then continues as a good path to (a,b) is $\frac{a-b}{a+b}$; this then is the solution to the ballot problem.

We return now to Catalan numbers. The k^{th} Catalan number c_k is expressible as ${}_2d_k$, the number of good paths from $(0,-1)$ to $(k,k-1)$. According to (1.1) this is

$$\binom{2k}{k} - \binom{2k}{k+1} = \frac{1}{k}\binom{2k}{k-1}.$$

Thus the k^{th} Catalan number c_k is given by the formula

$$c_0 = 1,\ c_k = \frac{1}{k}\binom{2k}{k-1},\ k \geqslant 1. \qquad (1.2)$$

Notice that an alternative expression is

$$c_k = \frac{1}{k+1}\binom{2k}{k},\ k \geqslant 0. \qquad (1.3)$$

Notice, too, that, by translation, c_k may also be interpreted as the number of good paths from (1,0) to $(k+1,k)$.

André's Reflection Method does not seem to be readily applicable to obtaining a formula corresponding to (1.2) in the case of a general $p \geqslant 2$. Indeed, in the next section, where we calculate ${}_pc_k$, $p \geqslant 2$, we do *not* obtain a general closed formula for the number of p-good paths from (c,d) to (a,b). We do introduce there and calculate explicitly the number d_{qk} of p-good paths from $(1,q-1)$ to $P_k(k,(p-1)k-1)$ for $q \leqslant p-1$. This enables us, by translation, to calculate the number of p-good paths from any lattice point C below the line $y = (p-1)x$ to P_k. For if $C = (c,d)$, then the number of p-good paths from C to P_k is the number of p-good paths from $(1,q-1)$ to P_{k-c+1}, where $q = d+1-(p-1)(c-1)$. On the other hand, although a convenient explicit formula when the terminus of the path is an arbitrary lattice point below the line $y = (p-1)x$ does not seem available, we will nevertheless derive in Section 3 an *algorithmic formula* in the form of a finite sum, depending on the quantities d_{qk}. This formula, while readily applicable, has a mysterious conceptual feature to which we draw attention at the end of Section 3.

The standard procedure for calculating c_k (we here revert to the original case $p = 2$) is based on the evident recurrence relation

$$c_k = \sum_{i+j=k-1} c_i c_j,\ k \geq 1;\ c_0 = 1. \tag{1.4}$$

This relation is most easily seen by considering b_k and concentrating on the *last* application of a given binary operation within a certain expression. If we form the power series

$$S(x) = \sum_{k=0}^{\infty} c_k x^k, \tag{1.5}$$

then (1.4) shows that S satisfies

$$xS^2 - S + 1 = 0,\ S(0) = 1,$$

so that

$$S = \frac{1 - \sqrt{1-4x}}{2x}. \tag{1.6}$$

It is now straightforward to expand (1.6) and thus to obtain confirmation of the value of c_k already obtained in (1.2). As we will demonstrate, this analytical method for calculating c_k does generalize, but the generalization is highly sophisticated and we prefer to emphasize a much more elementary, combinatorial argument to calculate ${}_pc_k$.

2. Generalized Catalan Numbers

We will base our calculation of ${}_pc_k$ on that of a more general quantity. Let p be a fixed integer greater than 1, and let P_k be the point $(k,(p-1)k-1)$, $k \geq 0$. We define the numbers ${}_pd_{qk} = d_{qk}$, as follows.

Definition 2.1 Let $q \leq p-1$. Then $d_{q0} = 1$ and d_{qk} is the number of p-good paths from $(1, q-1)$ to P_k, if $k \geq 1$.

The quantities d_{qk} will play a crucial role in counting p-good paths and turn out to be no more difficult to calculate than the quantities d_k. Of course, we have

$${}_pc_k = {}_pd_k = {}_pd_{0k},\ k \geq 0, \tag{2.1}$$

where ${}_pc_k$ is the k^{th} (generalized) Catalan number. For, if $k \geq 1$, every p-good path from $(0,-1)$ to P_k must first proceed to $(1,-1)$. We further note that we have the relation

$$d_{p-1,k} = d_{0,k-1},\ k \geq 1; \tag{2.2}$$

for an easy translation argument shows that the number of p-good paths from $(1, p-2)$ to P_k is the same as the number of p-good paths from $(0,-1)$ to P_{k-1}. Thus $d_{p-1,1} = 1 = d_{00}$ if $k = 1$, and, if $k \geq 2$, the number of p-good paths from $(0,-1)$ to P_{k-1} is, as argued above, the number of p-good paths from $(1,-1)$ to P_{k-1}, that is, $d_{0,k-1}$.

We may write ${}_pd_k$ for ${}_pd_{0k}$, so that, if $k \geq 0$,

$${}_2d_k = c_k, \text{ the } k^{\text{th}} \text{ Catalan number, and} \tag{2.3}$$

$${}_pc_k = {}_pd_k = {}_pd_{0,k} = {}_pd_{p-1,k+1}, \text{ the generalized } k^{\text{th}} \text{ Catalan number.} \tag{2.4}$$

We will also suppress the 'p' from these symbols and from the term 'p-good' if no ambiguity need be feared.

We now enunciate two fundamental properties of the numbers d_{qk}, for a fixed $p \geq 2$. A special case of the first property was pointed out to us by David Jonah during a Mathematical Association of America Short Course, conducted by us in the summer of 1987, under the auspices of the Michigan Section. His original formula concerned the case $p = 2$ and was restricted to $q = 0$ (and the parameter n appearing in the formula was restricted to being an integer not less than $2k$ instead of an arbitrary real number).

Recall that, for any *real* number n, the binomial coefficient $\binom{n}{0}$ may be interpreted as 1, and the binomial coefficient $\binom{n}{r}$ may be interpreted as the expression $\frac{n(n-1)\cdots(n-r+1)}{r!}$, provided r is a positive integer (see, e.g., [3]). Using this interpretation, we prove the following.

THEOREM 2.2 (Generalized Jonah Formula). *Let n be any real number. If $k \geq 1$, and $q \leq p-1$, then*

$$\binom{n-q}{k-1} = \sum_{i=1}^{k} d_{qi} \binom{n-pi}{k-i}. \tag{2.5}$$

Proof: We first assume (see Figure 9) that n is an integer not less than pk, so that $(k, n-k)$ is a lattice point not below the line $y = (p-1)x$. Partition the paths from $(1, q-1)$ to $(k, n-k)$ according to where they first meet the line $y = (p-1)x$. The number of these paths that first meet this line where $x = i$ is $\gamma_i \delta_i$, where

γ_i = the number of paths from $(1, q-1)$ to $(i, (p-1)i)$ which stay below $y = (p-1)x$ except at the endpoint
= the number of good paths from $(1, q-1)$ to P_i
= d_{qi};

and

δ_i = the number of paths from $(i, (p-1)i)$ to $(k, n-k)$
= $\binom{n-pi}{k-i}$.

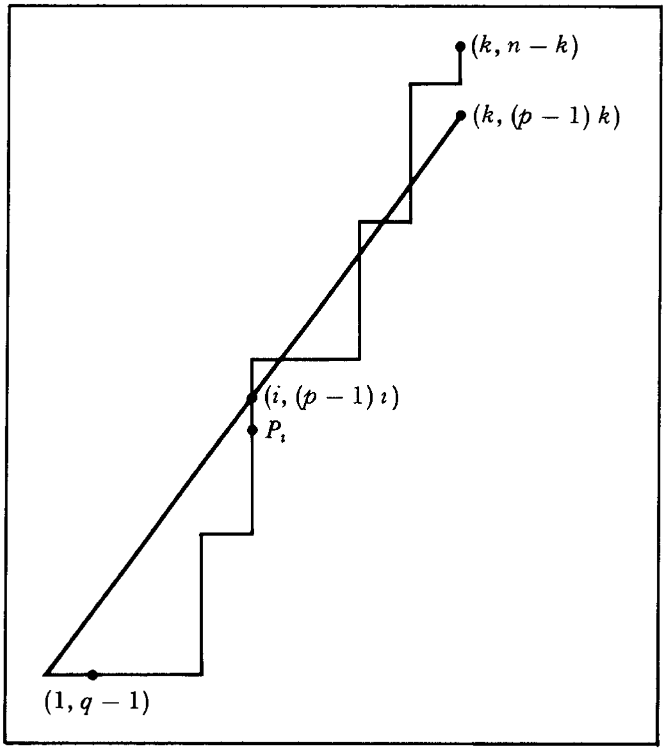

Figure 9. Partitioning paths from $(1, q-1)$ to $(k, n-k)$ according to where they first meet the line $y = (p-1)x$.

Because i ranges from 1 to k and there are $\binom{n-q}{k-1}$ paths in all from $(1, q-1)$ to $(k, n-k)$, formula (2.5) is proved in this case.

We now observe that each side of (2.5) is a polynomial in n over the rationals Q of degree $(k-1)$, and we have proved that these two polynomials agree for infinitely many values of n. They therefore agree for all *real* values of n.

We may use Theorem 2.2 to calculate the value of d_{qk}; recall that $q \leqslant p-1$.

THEOREM 2.3 *If* $k \geqslant 1$, *then*

$$d_{qk} = \frac{p-q}{pk-q}\binom{pk-q}{k-1}. \tag{2.6}$$

Proof: Because, from its definition, $d_{q1} = 1$, formula (2.6) holds if $k = 1$. We therefore assume $k \geqslant 2$. We first substitute[3] $n = pk - 1$ into (2.5), obtaining

$$\binom{pk-q-1}{k-1} = \sum_{i=1}^{k} d_{qi}\binom{pk-pi-1}{k-1}.$$

We next substitute $n = pk - 1$ in (2.5), but now replace k by $(k-1)$, obtaining

$$\binom{pk-q-1}{k-2} = \sum_{i=1}^{k-1} d_{qi}\binom{pk-pi-1}{k-i-1}.$$

[3] It is amusing to note that the *combinatorial* argument in the proof of Theorem 2.2 required $n \geqslant pk$, and we are here substituting a value of n less than pk.

(Recall that $k \geqslant 2$.)

Now we have the universal identity $\binom{n}{r+1} = \frac{n-r}{r+1}\binom{n}{r}$. Hence

$$\binom{pk-pi-1}{k-1} = \frac{pk-pi-k+i}{k-i}\binom{pk-pi-1}{k-i-1}$$

$$= (p-1)\binom{pk-pi-1}{k-i-1},$$

so that

$$\binom{pk-q-1}{k-1} - (p-1)\binom{pk-q-1}{k-2}$$

$$= d_{qk}\binom{-1}{0} = d_{qk}.$$

Finally,

$$\binom{pk-q-1}{k-1} - (p-1)\binom{pk-q-1}{k-2}$$

$$= \binom{pk-q}{k-1}\left\{\frac{pk-q-k+1}{pk-q} - \frac{(p-1)(k-1)}{pk-q}\right\}$$

$$= \frac{p-q}{pk-q}\binom{pk-q}{k-1},$$

and (2.6) is proved.

Note that nothing prevents us from taking q negative in Theorems 2.2 and 2.3. Taking $q = 0$, we obtain the values of the generalized Catalan numbers:

Corollary 2.4 ${}_pd_0 = 1$ *and* ${}_pd_k = \frac{1}{k}\binom{pk}{k-1}$, $k \geqslant 1$.

We come now to the second fundamental property of the numbers d_{qk}, for a fixed $p \geqslant 2$.

THEOREM 2.5 (Recurrence relation for d_{qk}). *Choose a fixed* $r \geqslant 1$. *If* $k \geqslant 1$, *and* $q < p - r$, *then*

$$d_{qk} = \sum_{i,j} d_{p-r,i}d_{q+r,j}, \text{ where } i \geqslant 1, j \geqslant 1, i+j = k+1. \tag{2.7}$$

The proof proceeds by partitioning the good paths from $(1, q-1)$ to P_k according to where they first meet the line $y = (p-1)x - r$ (see Figure 10). We suppress the details of the argument, which resemble those for Theorem 2.2. However, notice that formula (2.7) generalizes the familiar recurrence relation (1.4),

$$c_k = \sum_{i+j=k-1} c_i c_j,\ k \geqslant 1,$$

for the Catalan numbers, in the light of (2.3) and (2.4) —we have merely to substitute $p = 2$, $q = 0$, $r = 1$. In

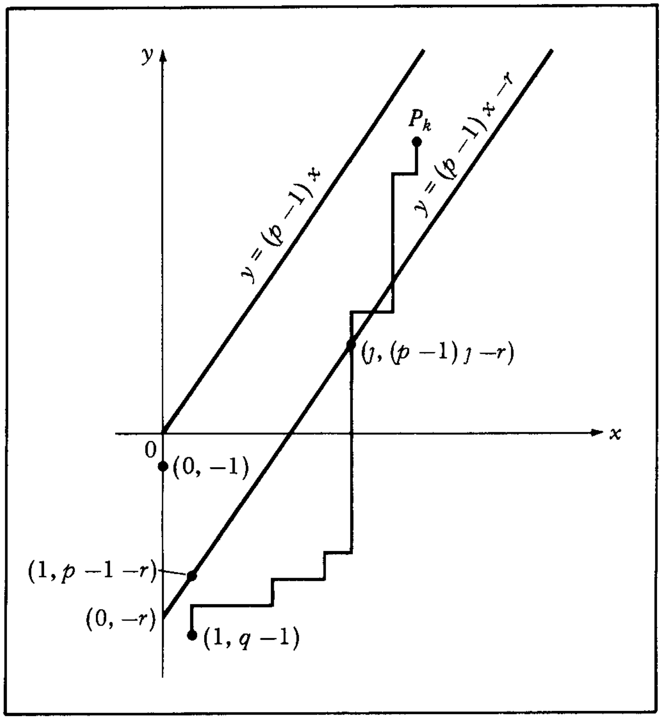

Figure 10. Proof of Theorem 2.5.

fact, we may use Theorem 2.5 to express d_{qk} in terms of the generalized Catalan numbers ${}_pd_i$, at the same time generalizing (2.4):

THEOREM 2.6. *If $q \leqslant p - 1$, and if $k \geqslant 1$, then*

$$d_{qk} = \sum_{i_1+i_2+\cdots+i_{p-q}=k-1} {}_pd_{i_1}\,{}_pd_{i_2}\cdots{}_pd_{i_{p-q}}. \qquad (2.8)$$

Proof: The case $q = p - 1$ is just (2.4), so we assume $q < p - 1$. Then, by Theorem 2.5, with $r = 1$,

$$d_{qk} = \sum_{i,j} d_{p-1,i}\, d_{q+1,j}, \quad \text{where } i \geqslant 1,\ j \geqslant 1,\ i + j = k + 1$$

$$= \sum_{i_1,j_2} {}_pd_{i_1} d_{q+1,j_2}, \text{ where } j_2 \geqslant 1,\ i_1 + j_2 = k, \quad \text{using (1.4).}$$

Iterating this formula, we obtain

$$d_{qk} = \sum_{i_1+i_2+\cdots+i_{p-q-1}+j_{p-q}=k} {}_pd_{i_1}\,{}_pd_{i_2}\cdots{}_pd_{i_{p-q-1}}\,{}_pd_{p-1,j_{p-q}}, \quad \text{where } j_{p-q} \geqslant 1.$$

A second application of (2.4) yields the formula (2.8).

Setting $q = 0$ in (2.8), and again using (2.4), yields our next result.

Corollary 2.7 If $k \geqslant 1$, then

$${}_pd_k = \sum_{i_1+i_2+\ldots+i_p=k-1} {}_pd_{i_1}\,{}_pd_{i_2}\cdots{}_pd_{i_p}.$$

Because the numbers ${}_pb_k$ are easily seen to satisfy the same relation, Corollary 2.7 provides another proof of the equality of ${}_pd_k$ with the k^{th} generalized Catalan number. It also shows us that, if $S(x)$ is the power series

$$S(x) = \sum_{k=0}^{\infty} {}_pd_k x^k,$$

then

$$xS^p = S - 1. \qquad (2.9)$$

Klarner [6] attributes to Pólya-Szegö [7] the observation that we may invert (2.9) to obtain

$$S(x) = 1 + \sum_{k=1}^{\infty} \frac{1}{k}\binom{pk}{k-1} x^k;$$

indeed, this is the solution given to Problem 211 on p. 125 of [7], based on the theory of Bürmann-Lagrange series. Thus, Corollary 2.7 leads to a different, but far more sophisticated, evaluation of the generalized Catalan numbers ${}_pd_k$, though it does not, in any obvious way, yield Theorem 2.3.

Note that Corollary 2.4 could, of course, have been obtained without introducing the quantities ${}_pd_{qk}$ for $q \neq 0$. However, these quantities have an obvious combinatorial significance and satisfy a recurrence relation (Theorem 2.5) much simpler than that of Corollary 2.7. In the light of (2.1) the quantities ${}_pd_{qk}$ themselves deserve to be regarded as generalizations of the Catalan numbers—though certainly not the ultimate generalization! We will see in the next section that they are useful in certain important counting algorithms.

3. Counting p-good Paths

We will give in this section an algorithm for counting the number of p-good paths from (c,d) to (a,b), each of those lattice points being assumed below the line $y = (p - 1)x$. As explained in Section 1, it suffices (via a translation) to replace (c,d) by the point $(1, q - 1)$, with $q \leqslant p - 1$. Further, in order to blend our notation with that of Section 2, we write $(k, n - k)$ for (a,b), so that $n < pk$. We assume there are paths from $(1, q - 1)$ to $(k, n - k)$, i.e., that $k \geqslant 1$, $n - k \geqslant q - 1$.

Of course, there are $\binom{n-q}{k-1}$ paths from $(1, q - 1)$ to $(k, n - k)$. It is, moreover, easy to see that there exist p-bad paths from $(1, q - 1)$ to $(k, n - k)$ if and only if $n - k \geqslant p - 1$, so we assume this—otherwise, there are only p-good paths and we have our formula, namely, $\binom{n-q}{k-1}$.

To sum up, we are counting the p-good paths from $(1, q - 1)$ to $(k, n - k)$ under the (nontrivializing) assumption that $1 \leqslant k \leqslant n - p + 1 \leqslant p(k - 1)$; notice

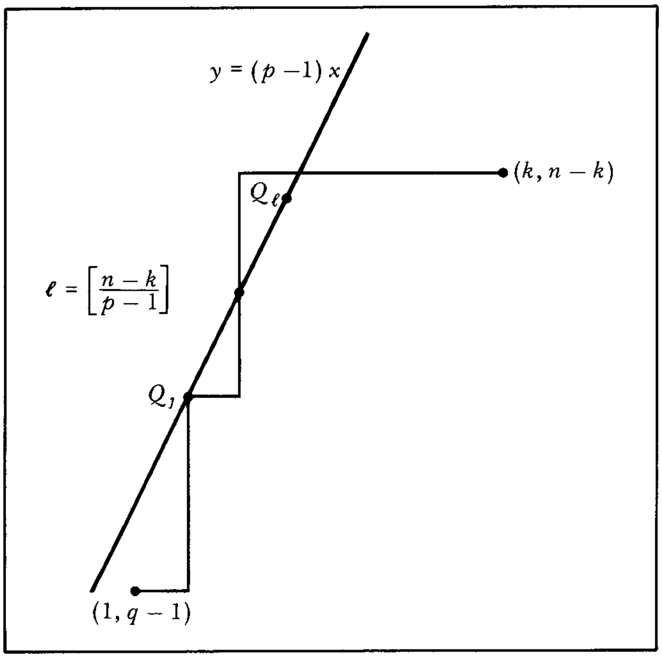

Figure 11. A bad path from $(1, q-1)$ to $(k, n-k)$, $n < pk$. Notice that Q_ℓ is the last possible initial crossing point of the line $y = (p-1)x$ for such a path.

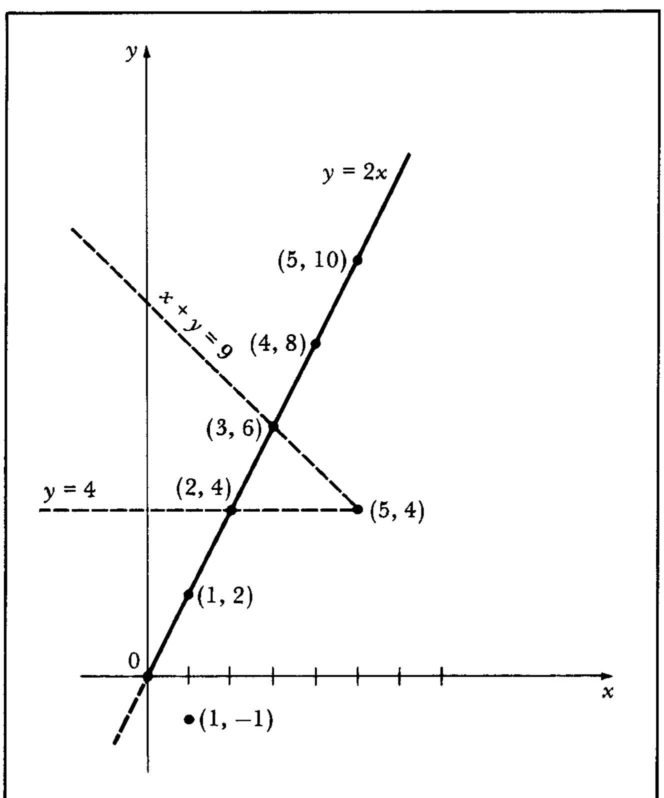

Figure 12. Corollary 3.3 with $q = 0$, $k = 5$, $p = 3$, $n = 9$; hence $l = 2$, $m = 3$.

$$\overbrace{\binom{9}{4}}^{\text{all paths}} = \overbrace{{}_3d_1\binom{6}{4} + {}_3d_2\binom{3}{3}}^{\text{bad paths}} + {}_3d_3\binom{0}{2} + \overbrace{{}_3d_4\binom{-3}{1} + {}_3d_5\binom{-6}{0}}^{\text{good paths}}$$
$$= 1(15) + 3(1) + 12(0) + 55(-3) + 273(1)$$
$$\text{i.e., } 126 = 18 + 108$$

that this composite inequality implies that, in fact, $k \geq 2$, so that we assume

$$2 \leq k \leq n - p + 1 \leq p(k-1). \tag{3.1}$$

There are $\binom{n-q}{k-1}$ paths in all, so we count the p-bad paths. Let $\ell = [\frac{n-k}{p-1}]$, where $[x]$ is the integral part of x. Then (3.1) ensures that $\ell \geq 1$, and a p-bad path from $(1, q-1)$ to $(k, n-k)$ will meet the line $y = (p-1)x$ first at some point $Q_j(j,(p-1)j)$, $j = 1, 2, \ldots, \ell$ (see Figure 11).

Because there are d_{qj} paths from $(1, q-1)$ to Q_j that stay below the line $y = (p-1)x$ except at Q_j, and because there are $\binom{n-pj}{k-j}$ paths from Q_j to $(k, n-k)$, it follows that there are $d_{qj}\binom{n-pj}{k-j}$ bad paths from $(1, q-1)$ to $(k, n-k)$ that first meet the line $y = (p-1)x$ at Q_j. Thus the number of p-bad paths from $(1, q-1)$ to $(k, n-k)$ is

$$\sum_{j=1}^{\ell} d_{qj}\binom{n-pj}{k-j}.$$

We have proved the following result.

THEOREM 3.1. *The number of p-good paths from $(1, q-1)$ to $(k, n-k)$, under the conditions* (3.1), *is*

$$\binom{n-q}{k-1} - \sum_{j=1}^{\ell} d_{qj}\binom{n-pj}{k-j}, \text{ where } \ell = \left[\frac{n-k}{p-1}\right].$$

We may express this differently, however, and in a way that will be more convenient if ℓ is relatively large. If we invoke Theorem 2.2, we obtain the following corollary.

Corollary 3.2 The number of p-good paths from $(1, q-1)$ to $(k, n-k)$, under the conditions (3.1), *is*

$$\sum_{j=\ell+1}^{k} d_{qj}\binom{n-pj}{k-j}, \text{ where } \ell = \left[\frac{n-k}{p-1}\right].$$

It is interesting to observe that the binomial coefficients entering into the formula of Corollary 3.2 are certainly 'generalized,' because $k - j > n - pj$ if $j > \ell$. Thus if $\ell + 1 \leq j \leq k$, then $\binom{n-pj}{k-j} = 0$, so long as $n - pj \geq 0$, i.e., $j \leq [n/p]$. Because $n/p > (n-k)/(p-1)$, it follows that $[n/p] \leq [(n-k)/(p-1)]$, so that we may improve Corollary 3.2, at least formally, by replacing ℓ by $m = [n/p]$. Thus we obtain our final corollary.

Corollary 3.3. The number of p-good paths from $(1, q-1)$ to $(k, n-k)$, under the conditions (3.1), *is*

$$\sum_{j=m+1}^{k} d_{qj}\binom{n-pj}{k-j}, \text{ where } m = \left[\frac{n}{p}\right].$$

We give an example (see Figure 12).

Example 3.1. Let $q = 0$, $k = 5$, $p = 3$, $n = 9$. The inequalities (3.1) are certainly satisfied and $m = 3$. Thus the number of 3-good paths from $(1,-1)$ to $(5,4)$ is

$$d_4\binom{-3}{1} + d_5\binom{-6}{0} = -3d_4 + d_5.$$

Now d_j is the j^{th} (generalized) Catalan number for $p = 3$; thus, by Corollary 2.4, $d_4 = \frac{1}{4}\binom{12}{3} = 55$ and $d_5 = \frac{1}{5}\binom{15}{4} = 273$. Thus the number of 3-good paths is 108. Notice that the total number of paths is $\binom{9}{4} = 126$. Notice also that, in this case, $\ell = [(n-k)/(p-1)] = 2$, so that there is an advantage in replacing ℓ by m.

Circumstances will determine whether it is more convenient to use Theorem 3.1 or Corollary 3.3. In the example above there is little to choose. However, whereas in Theorem 3.1 each term has an obvious combinatorial meaning, it is difficult to assign a combinatorial meaning to the terms in Corollary 3.3; the binomial coefficient $\binom{n-pj}{k-j}$, $j \geq m+1$, does not count the paths from Q_j to $(k, n-k)$, because there are none —might it be said to count 'phantom paths'?

Example 3.1 is really "generic" and thus merits closer study. Theorem 2.2 tells us, in this case, that

$$\binom{9}{4} = \sum_{i=1}^{5} d_i\binom{9-3i}{5-i}. \tag{3.2}$$

The expression of the right of (3.2) breaks up into

$$\underbrace{d_1\binom{6}{4} + d_2\binom{3}{3}} + \underbrace{d_3\binom{0}{2}} + \underbrace{d_4\binom{-3}{1} + d_5\binom{-6}{0}}. \tag{3.3}$$

Now the binomial coefficient $\binom{9}{4}$ of the left of (3.2) counts the number of paths from $(1,-1)$ to $(5,4)$; the first block in (3.3) counts the bad paths from $(1,-1)$ to $(5,4)$, partitioning them into those that first meet the danger line $y = 2x$ at $(1,2)$ and those that first meet the danger line at $(2,4)$; the second block in (3.3) makes a zero contribution; and the third block in (3.3) thus counts the good paths from $(1,-1)$ to $(5,4)$. In some mysterious sense the third block—which intrigues us enormously—is also partitioned by the points $(4,8)$ and $(5,10)$ on the line $y = 2x$. But this is only a notational partitioning, because we have 108 represented as $(55)(-3) + (273)(1)$. Here the coefficients $d_i (i = 4,5)$ have an obvious combinatorial significance, but the binomial coefficients, $\binom{-3}{1}$ and $\binom{-6}{0}$, do not. They result from applying the formula $\binom{n-pi}{k-i}$ for the number of paths from $(i, (p-1)i)$ to $(k, n-k)$ *outside* its domain of validity $1 \leq i \leq \ell$. There is here much food for thought!

Historical Note

The versatile Belgian mathematician Eugène Charles Catalan (1814–1894) defined the numbers named after him in connection with his solution of the problem of dissecting a polygon by means of non-intersecting diagonals into triangles; thus his definition of the k^{th} Catalan number is our c_k. In fact, the problem of enumerating such dissections had already been solved by Segner in the eighteenth century, but not in convenient or perspicuous form. Euler (see the bibliography) had immediately obtained a simpler solution involving generating functions—we have sketched this approach in Section 1. Catalan's own solution, achieved almost simultaneously with that of Binet around 1838, was even simpler, not requiring the theory of generating functions, which was regarded at the time as rather 'delicate' in view of its (apparent) dependence on the notion of convergence of series.

A host of other interpretations of the Catalan numbers have been given, largely inspired by work in combinatorics and graph theory, and have been collected by Gould (see the bibliography). The interpretations have led to generalizations, of which the most obvious and important, treated in this article, have been studied by probabilists (e.g., Feller) and others. The ballot problem itself leads to an evident generalization, considered by Feller, in which the line $y = x$ is replaced by the line $y = (p-1)x$.

An interesting generalization of a somewhat different kind is treated by Wenchang Chu (see the bibliography).

References

1. D. André, Solution directe du problème résolu par M. Bertrand, *Comptes Rendus Acad. Sci. Paris* 105 (1887), 436–7.
2. A. P. Hillman, G. L. Alexanderson and R. M. Grassl, *Discrete and Combinatorial Mathematics*, San Francisco: Dellen (1987).
3. Peter Hilton and Jean Pedersen, Extending the binomial coefficients to preserve symmetry and pattern, *Computers Math. Applic.* 17 (1–3) (1989), 89–102.
4. Peter Hilton and Jean Pedersen, The ballot problem and Catalan numbers, *Nieuw Archief voor Wiskunde*, 1990 (to appear).
5. A. Hurwitz and R. Courant, *Allgemeine Funktionentheorie und elliptische Funktionen. Geometrische Funktionentheorie*, Berlin: Springer (1922), 128.
6. David A. Klarner, Correspondences between plane trees and binary sequences, *J. of Comb. Theory* 9 (1970), 401–411.
7. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Vol. I, Berlin, Göttingen, Heidelberg: Springer (1954), 125.
8. N. J. A. Sloane, *A Handbook of Integer Sequences*, New York and London: Academic Press (1973).
9. Marta Sved, Counting and recounting, *Mathematical Intelligencer* 5, no. 4 (1983), 21–26.

Bibliography

[Items [2], [6] and [8] among the References may also be regarded as general sources on Catalan numbers.]

R. Alter, Some remarks and results on Catalan numbers, *Proc. 2nd Louisiana Conf. on Comb., Graph Theory and Comp.* (1971), 109–132.

D. André, Solution directe du problème résolu par M. Bertrand, *Comptes Rendus Acad. Sci. Paris* 105 (1887), 436–7.

W. G. Brown, Historical note on a recurrent combinatorial problem, *Amer. Math Monthly* 72 (1965), 973–977.

Wenchang Chu, A new combinatorial interpretation for generalized Catalan numbers, *Discrete Math.* 65 (1987), 91–94.

K. L. Chung and W. Feller, On fluctuations in coin-tossing, *Proc. Nat. Acad. Sci., U.S.A.* 35 (1949), 605–608.

John H. Conway and Richard K. Guy, *The Book of Numbers*, Sci. Amer. Library, W. H. Freeman (to appear 1990).

I. Z. Chorneyko and S. G. Mohanty, On the enumeration of certain sets of planted trees, *J. Combin. Theory Ser.* B18 (1975), 209–21.

T. T. Cong and M. Sato, One-dimensional random walk with unequal step lengths restricted by an absorbing barrier, *Discrete Math.* 40 (1982), 153–162.

R. Donoghey and L. W. Shapiro, Motzkin numbers, *J. Combin. Theory Ser.* A23 (1977), 291–301.

N. Dershowitz and S. Zaks, Enumeration of ordered trees, *Discrete Math.* 31 (1980), 9–28.

L. Euler, *Novi Commentarii Academiae Scientarium Imperialis Petropolitanque* 7 (1758/9), 13–14.

Roger B. Eggleton and Richard K. Guy, Catalan strikes again! How likely is a function to be convex?, *Math. Mag.* 61 (1988), 211–219.

A. Erdélyi and I. M. H. Etherington, Some problems of non-associative combinatorics, *Edinburgh Math. Notes.* 32 (1941), 7–12.

I. J. Good, Generalizations to several variables of Lagrange's expansion, *Proc. Camb. Phil. Soc.* 56 (1960), 367–380.

Henry W. Gould, *Bell and Catalan Numbers, Research bibliography of two special number sequences*, available from the author (Department of Mathematics, West Virginia University, Morgantown, WV 26506). The 1979 edition sells for $3.00 and contains over 500 references pertaining to Catalan numbers.

Henry W. Gould, *Combinatorial Identities*, available from the author (Department of Mathematics, West Virginia University, Morgantown, WV 26506).

Richard K. Guy, Dissecting a polygon into triangles, *Bulletin of the Malayan Mathematical Society* 5 (45) (1958), 57–60.

Richard K. Guy, A medley of Malayan mathematical memories and mantissae, *Math. Medley* (Singapore), 12, no. 1 (1984), 9–17.

V. E. Hoggatt, Jr., and Marjorie Bicknell, Pascal, Catalan, and general sequence convolution arrays in a matrix, *Fibonacci Quarterly* 14, no. 2, 135–143.

V. E. Hoggatt, Jr., and Marjorie Bicknell, Sequences of matrix inverses from Pascal, Catalan, and related convolution arrays, *Fibonacci Quarterly* 14, no. 3, 224–232.

V. E. Hoggatt, Jr., and Marjorie Bicknell, Catalan and related sequences arising from inverses of Pascal's triangle matrices, *Fibonacci Quarterly* (December 1976), 395–405.

M. S. Klamkin, Problem 4983, *Amer. Math. Monthly* 69 (1962), 930–931.

Mike Kuchinski, *Catalan structures and Correspondences*, M.Sc. thesis, Department of Mathematics, West Virginia University, Morgantown, WV 26506.

Th. Motzkin, Relations between hypersurface cross ratios, and a combinatorial formula for partitions of a polygon, and for non-associative products, *Bull. Amer. Math. Soc.* 54 (1948), 352–360.

G. Pólya, On picture-writing, *Amer. Math. Monthly* 63 (1956), 689–697.

G. N. Raney, Functional composition patterns and power series inversion, *Trans. Amer. Math. Soc.* 94 (1960), 441–451.

D. G. Rogers, Pascal triangles, Catalan numbers and renewal arrays, *Discrete Math.* 22 (1978), 301–310.

A. D. Sands, On generalized Catalan numbers, *Discrete Math.* 21 (1978), 218–221.

Memoirs by Eugène-Charles Catalan relevant to the theme of Catalan numbers may be found in *Journal de Mathématiques pures et appliquées de Liouville*, (1), III, 508–816; IV, 91–94, 95–99; VI, 74, 1838–41.

For details of Catalan's life and work see "Les travaux mathématiques de Eugène-Charles Catalan," *Annuaire de L'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, Brussels (1896), 115–172.

Peter Hilton
Department of Mathematical Sciences
State University of New York
Binghamton, NY 13901 USA
and
Abteilung für Mathematik
E.T.H., 8092 Zürich, Switzerland

Jean Pedersen
Department of Mathematics
Santa Clara University
Santa Clara, CA 95053 USA