



## A Combinatorial Interpretation of a Catalan Numbers Identity

David Callan

*Mathematics Magazine*, Vol. 72, No. 4. (Oct., 1999), pp. 295-298.

Stable URL:

<http://links.jstor.org/sici?sici=0025-570X%28199910%2972%3A4%3C295%3AACIOAC%3E2.0.CO%3B2-J>

*Mathematics Magazine* is currently published by Mathematical Association of America.

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/maa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

---

# NOTES

---

## A Combinatorial Interpretation of a Catalan Numbers Identity

DAVID CALLAN  
 Department of Statistics  
 University of Wisconsin-Madison  
 1210 W. Dayton Street  
 Madison, WI 53706-1693

**Introduction** The Catalan numbers  $C_n := \frac{1}{n+1} \binom{2n}{n} = \binom{2n}{n} - \binom{2n}{n-1}$  seem to be ubiquitous in combinatorics, and they have featured frequently in these pages. A list of some fifteen more or less different situations in which they occur is given in [2], and they make an unexpected appearance in a problem on the expected length of a World Series [3].

Their generating function  $\sum_{n=0}^{\infty} C_n x^n$  is given by  $C(x) := \frac{1 - \sqrt{1-4x}}{2x}$  (expand the radical by the binomial theorem). Once an identity involving  $C(x)$  is found—by whatever means—the question of a combinatorial interpretation arises. For example, isolating the radical and squaring leads to a fundamental identity:  $x C(x)^2 = C(x) - 1$ . Equating coefficients of  $x$ , this says that  $C_n = C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-1} C_0$ ,  $n \geq 1$ , an identity with a well-known interpretation: the left side  $C_n$  counts Catalan paths of  $2n$  steps (see below) and the right side counts these paths by “first return to the diagonal.”

The following identity is obtained in [4] in connection with “moments of a Catalan triangle”:

$$\sum_{k \geq 1} k(C(x) - 1)^k = \sum_{n \geq 1} 4^{n-1} x^n. \quad (1)$$

(Note that  $C(x) - 1$  has constant term 0 and hence the left side of (1) makes sense.) The object of this note is to give a combinatorial interpretation of (1) using lattice paths; that is, we describe two sets of such paths whose cardinalities are the coefficients of  $x^n$  on the two sides of (1), and then give a bijection between these two sets.

**Combinatorial setting** To set the stage, let  $E$  and  $N$  denote unit steps East and North respectively. A *path* is (for present purposes) a finite sequence of an *even* number of steps and it can be represented graphically in an obvious way. For example,  $ENEENE$  is given in FIGURE 1 with its steps numbered.

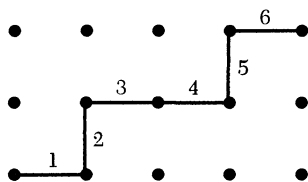


FIGURE 1

It turns out to be helpful to focus on the consecutive *pairs* of steps comprising a path (while not forgetting entirely about the underlying individual steps). Let us denote the pairs  $EE$ ,  $NN$ ,  $EN$ ,  $NE$  by  $H$  (for horizontal),  $V$  (vertical),  $L$  (lower hook),  $U$  (upper hook) respectively, and call them *double-steps* for short. Thus the path in FIGURE 1 above is also specified by  $LHU$ . A *subpath* of a path  $P$  is any subsequence of consecutive double-steps in the sequence of double-steps that defines  $P$ . In other words, a subpath must contain an even number of steps and begin at an odd-numbered step. Thus, in FIGURE 1, steps numbered 34 and 3456 form subpaths, but 123, 1256, and 2345 do not.

If a path has  $2n$  steps, and hence  $n$  double-steps, its *length* is  $n$ , and we call it an  $n$ -*path*. Let  $\mathcal{P}_n$  denote the set of  $n$ -paths. Since the steps of a path can be chosen independently,  $|\mathcal{P}_n| = 2^{2n} = 4^n$  and so  $|\mathcal{P}_{n-1}|$  is the coefficient of  $x^n$  on the right in (1).

To bring Catalan numbers into the picture, define a *Catalan path* to be a path containing an equal number of  $E$ 's and  $N$ 's, whose graph lies entirely on or below the line—necessarily of slope 1—joining its initial and terminal points. We'll call this the *diagonal line* of the path. The two Catalan 2-paths are given in FIGURE 2.

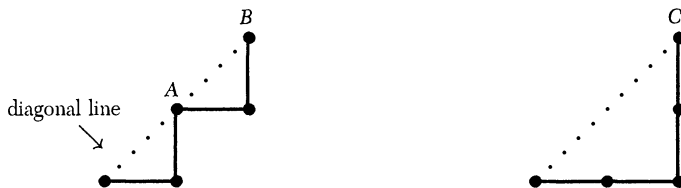


FIGURE 2

In fact the number of Catalan  $n$ -paths is the Catalan number  $C_n$  (see, for example, [1]).

Catalan paths can be strung together by concatenation, maintaining catalan-icity, so it is useful to identify the points where a Catalan path  $P$  returns to its diagonal line. Any such point (with the exception of the initial point) is called a *return* of  $P$ . Thus the Catalan paths in FIGURE 2 have two returns (at  $A$  and  $B$ ) and only one return (at  $C$ ) respectively. (The terminal point is always a return.) We use  $[x^n]R(x)$  to denote the coefficient of  $x^n$  in  $R(x)$  and our next task is to find some class of combinatorial objects counted by  $[x^n]\sum_{k \geq 1} k(C(x) - 1)^k$ . This is not as hard as it may seem. First, by direct expansion,  $[x^n](C(x) - 1)^k$  is  $\sum C_{i_1} C_{i_2} \cdots C_{i_k}$ , where the sum runs over all  $k$ -tuples  $(i_1, i_2, \dots, i_k)$  of *positive* integers whose sum is  $n$ , and hence is the number of  $k$ -tuples of *nonvanishing* Catalan paths whose total length is  $n$ . Imagine marking the terminal points of these  $k$  paths and concatenating them to form a single object, a “marked” Catalan path.

Thus we define a  $k$ -*marked Catalan path* as one in which  $k$  of its return points (arbitrary except that they must include its terminal point) have been distinguished by “marking” them. Since the original  $k$ -tuple of paths can be uniquely retrieved from the single  $k$ -marked path, we see that  $[x^n](C(x) - 1)^k$  is the number of  $k$ -marked Catalan  $n$ -paths. Now to account for the factor  $k$  on the left in (1), we need  $k$  copies of each  $k$ -marked path,  $k \geq 1$ . An obvious way to distinguish between  $k$  such copies is to “highlight” in each a different one of its  $k$  marks. Let  $\mathcal{H}_n$  denote the set of all highlighted Catalan  $n$ -paths (any number  $k \geq 1$  of marked returns). By construction of  $\mathcal{H}_n$ , therefore, the coefficient of  $x^n$  on the left side of (1) is  $|\mathcal{H}_n|$ .

**The bijection** Now we would like a bijection  $\pi: \mathcal{P}_{n-1} \rightarrow \mathcal{H}_n$ . Here is one. First note that if two Catalan subpaths of a given path  $P \in \mathcal{P}_{n-1}$  overlap (even just at an endpoint), their union is again a Catalan subpath because, unless one path is contained in the other, their diagonal lines must coincide. Hence the maximal Catalan subpaths of  $P_{n-1}$  are disjoint. Find these subpaths and leave them intact, but replace every other double-step by an upper hook, add an  $E$  at the start and an  $N$  at the finish. Denote the resulting path  $\bar{P}$ . An example is given in FIGURE 3, below. (The heavy lines represent the maximal Catalan subpaths of  $P$ .)

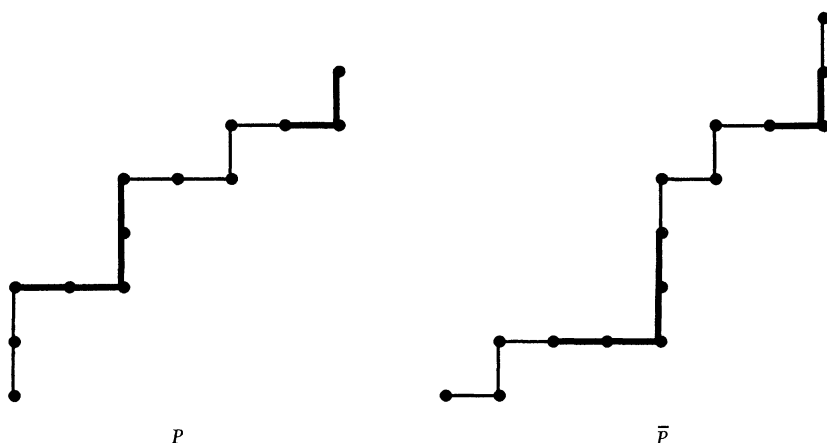


FIGURE 3

Clearly, in  $\bar{P}$  the diagonal lines of all Catalan paths left intact from  $P$  coincide, and  $\bar{P}$  is Catalan with diagonal line one unit higher. Next we will mark and highlight  $\bar{P}$  appropriately. The terminal point of  $\bar{P}$  is of course a return that must be marked. Every other return of  $\bar{P}$  is the midpoint of a double-step in  $P$  that was not part of a maximal—hence of any—Catalan subpath of  $P$  (and *vice versa*). None of these double-steps was a lower hook (which would already be a Catalan path). Thus the nonterminal returns in  $\bar{P}$  correspond only to  $H$ 's,  $V$ 's or  $U$ 's in  $P$ . Now mark those returns that correspond to  $H$ 's or  $V$ 's and suppose  $A_1, A_2, \dots, A_s$  is a listing, in order, of these  $H$ 's and  $V$ 's. Can we canonically choose one of the marked returns in  $\bar{P}$  to highlight? Yes: the key observation is that a  $V$  will never follow an  $H$  in the list  $A_1, A_2, \dots, A_s$ . For, if it did,  $P$  would have a subpath of the form  $H***V$ , where the asterisks represent a (possibly empty) sequence of upper hooks and/or Catalan subpaths. But then  $H***V$  would be a Catalan subpath (draw a picture to see this) whereas by definition of this  $H$  and  $V$ , they do not lie on any Catalan subpath. So the sequence  $\{A_i\}$  looks like  $V \cdots VH \cdots H$ . This allows us to highlight the first marked return in  $\bar{P}$  corresponding to an  $H$  in this sequence (or the terminal point of  $\bar{P}$  if the sequence contains no  $H$ 's). Thus, for  $\bar{P}$  in FIGURE 3 above, returns 1, 2, 4 get marked, the sequence  $\{A_i\}$  is  $VH$ , and so return 2 gets highlighted (with a circle).

Is this mapping  $\pi$  reversible? Yes! Given, say, the highlighted marked Catalan path in FIGURE 4a, draw its diagonal line and a parallel line one unit below it (the dotted lines in FIGURE 4a). Discard the initial and terminal steps and leave what's below the lower dotted line intact (heavy lines in FIGURE 4). Modify each upper hook that lies between the dotted lines as follows: leave it intact unless its midpoint is marked, in which case change it to a  $V$  if it lies (strictly) to the left of the highlight and to an  $H$  otherwise. Thus the path in FIGURE 4a becomes the path in FIGURE 4b.

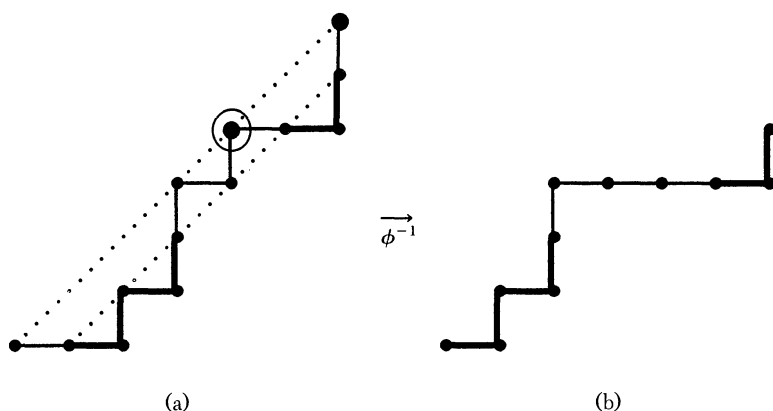


FIGURE 4

**Conclusion** Since  $\pi$  has an inverse, it is a bijection. To make this bijection fully explicit, an algorithm for identifying the maximal Catalan subpaths of a given path would be needed. As an interesting exercise, the reader might like to try to devise such an algorithm and perhaps write a computer program to implement it.

Finally, if you've read this far, you might like to try your hand at a simpler bijection: There are  $\binom{2n}{n}$  paths of  $n$  East and  $n$  North steps; hence  $\frac{1}{2}\binom{2n}{n}$  such paths that start with an East step. On the other hand, there are  $\frac{1}{2}\binom{2n}{n}$  marked Catalan  $n$ -paths (any number  $k \geq 1$  of marked returns, no highlight). See if you can find a bijection between these two sets. Hint: try flipping appropriate Catalan subpaths in the diagonal. If you succeed, you will have found a combinatorial interpretation for

$$\sum_{k=1}^{\infty} (C(x) - 1)^k = \frac{x}{2\sqrt{1-4x}},$$

an identity also given in [4].

**Acknowledgment.** Thanks to the referees for several helpful suggestions and reference [4].

## REFERENCES

1. D. Callan, Pair them up!: A visual approach to the Chung–Feller theorem, *College Math. J.* 26 (1995), 196–198
2. R. Eggleton and R. Guy, Catalan strikes again! How likely is a function to be convex?, this MAGAZINE 61 (1988), 211–219
3. L. Shapiro and W. Hamilton, The Catalan numbers visit the World Series, this MAGAZINE 66 (1993), 20–22
4. L. Shapiro, W.-J. Woan, S. Getu, Runs, slides and moments, *Siam J. Alg. Disc. Meth.* 4 (1983) 459–466