# The Catalan Numbers Re-Visit the World Series

Peter J. Larcombe,[†] Eric J. Fennessey,[‡] Wolfram A. Koepf[*]

and

David R. French (Retired)[†]

[†]School of Computing and Technology
University of Derby
North Tower
Kedleston Road, Derby DE22 1GB, U.K.
{P.J.Larcombe@derby.ac.uk}

[‡]Integrated Systems Division
AMS
Mountbatten House
Fitzherbert Road, Portsmouth PO6 1RU, U.K.
{eric.fennessey@amsjv.com}

[*]Department of Mathematics and Computer Science
University of Kassel
Heinrich-Plett-Strasse 40, 34132 Kassel, Germany
{koepf@mathematik.uni-kassel.de}

**Abstract**

An interesting statistical result, established in a 1993 paper by Shapiro and Hamilton in considering a certain type of theoretical tournament, is re-visited and leads to a new representation of the general Catalan number as a binomial coefficient sum.

# Introduction

Two teams compete in a first-to-$n$ games, World Series type, tournament (of maximum game duration $2n-1$) with, for each independent game, a respective constant win probability of $p$ and $q = 1-p$, where $0 < p, q < 1$; it is assumed that no game is drawn. In 1993 Shapiro and Hamilton [1] considered such a situation and, in particular, for prescribed $n > 0$ the expected number of games to be played before either team wins. Denoting this quantity as $E_n$, and the $(k+1)$th Catalan number as

$$c_k = \frac{1}{k+1}\binom{2k}{k}, \qquad k = 0, 1, 2, \ldots, \tag{1}$$

the following result was stated and proven.

**Theorem (Shapiro & Hamilton)** For $n \geq 1$,

$$\frac{E_n}{n} = \sum_{k=0}^{n-1} c_k (pq)^k.$$

This is an appealing and surprising form for $E_n/n$, since as a polynomial in $pq$ it acts as an ordinary generating function for finite $n$-subsequences of the (infinite) Catalan sequence $\{c_0, c_1, c_2, c_3, c_4, \ldots\} = \{1, 1, 2, 5, 14, \ldots\}$. Accordingly, in this paper the result is used to produce a new formulation of the general Catalan number. Our starting point is a different version of $E_n$, namely,

$$E_n = n \sum_{k=1}^{n} \binom{2n-k}{n-k} (pq)^{n-k}(p^k + q^k), \qquad n \geq 1, \tag{2}$$

which is accessible via undergraduate probability theory and requires, therefore, no supporting analysis (in fact it is proven formally in [1], in a mechanistic manner, *en route* to the above Theorem, but written with an incorrect lower limit $k = 0$). Equation (2) suggests that $p^k + q^k$ is expressible as a series in $pq$, a question that was the motivation for the work presented here. Writing $x = pq \in (0, \frac{1}{4}]$, and introducing the function

$$F_n(x) = p^n + q^n, \qquad n \geq 0, \tag{3}$$

scrutiny of a few low order cases confirms the answer as a positive one for these:

$$F_0(x) = p^0 + q^0 = 2,$$

$$
\begin{aligned}
F_1(x) &= p^1 + q^1 = 1, \\
F_2(x) &= p^2 + q^2 = (p+q)^2 - 2pq = 1 - 2x, \\
F_3(x) &= p^3 + q^3 = (p+q)^3 - 3pq(p+q) = 1 - 3x, \\
F_4(x) &= p^4 + q^4 = (p+q)^4 - 4pq(p^2+q^2) - 6p^2q^2 \\
&= 1 - 4xF_2(x) - 6x^2 \\
&= 1 - 4x + 2x^2, \\
F_5(x) &= p^5 + q^5 = (p+q)^5 - 5pq(p^3+q^3) - 10p^2q^2(p+q) \\
&= 1 - 5xF_3(x) - 10x^2 \\
&= 1 - 5x + 5x^2, \\
F_6(x) &= p^6 + q^6 = (p+q)^6 - 6pq(p^4+q^4) - 15p^2q^2(p^2+q^2) - 20p^3q^3 \\
&= 1 - 6xF_4(x) - 15x^2F_2(x) - 20x^3 \\
&= 1 - 6x + 9x^2 - 2x^3, \qquad (4)
\end{aligned}
$$

and so on. We find an appropriate polynomial form for $F_n(x)$, and from it establish a novel binomial coefficient sum representation of the Catalan number defined in (1). This is given in a later section, before which properties of the function—relevant to the formulation or considered to be of mathematical interest in their own right—are detailed.

# Properties of $F_n(x)$

First, we observe that $F_n(x)$ satisfies a simple linear second order recurrence equation (with starting values $F_0(x) = 2$, $F_1(x) = 1$).

**Lemma 1** For $n \geq 0$,

$$F_{n+2}(x) - F_{n+1}(x) + xF_n(x) = 0.$$

Proof Consider

$$
\begin{aligned}
F_{n+2}(x) &= p^{n+2} + q^{n+2} \\
&= p^{n+1}p + q^{n+1}q \\
&= p^{n+1}(1-q) + q^{n+1}(1-p) \\
&= p^{n+1} + q^{n+1} - pq(p^n + q^n) \\
&= F_{n+1}(x) - xF_n(x). \square
\end{aligned}
$$

Remark 1 The sequence $\{F_0(-1), F_1(-1), F_2(-1), \ldots\}$ is Fibonacci-like, the well known Fibonacci numbers $F_0^* = 0$, $F_1^* = 1$, $F_2^* = 1$, $F_3^* = 2$, $F_4^* = 3$, $F_5^* = 5$, $F_6^* = 8, \ldots$, say, being generated (given $F_0^*, F_1^*$) from the recursion $F_{n+2}^* - F_{n+1}^* - F_n^* = 0$ which corresponds in form to that above for $x = -1$ (probabilistically impossible, of course, in our tournament context).[1] The Fibonacci polynomials $F_0^*(x), F_1^*(x), F_2^*(x)$, *etc.*, satisfy $F_{n+2}^*(x) = xF_{n+1}^*(x) + F_n^*(x)$ (with $F_0^*(x), F_1^*(x)$ suitably defined), but cannot be related to our polynomials in a direct fashion; we shall, though, mention Fibonacci polynomials again (see just before Lemma 5).

**Lemma 2** The (ordinary) generating function for the sequence of polynomials $\{F_0(x), F_1(x), F_2(x), \ldots\}$ is

$$O(y;x) = \sum_{n=0}^{\infty} F_n(x)y^n = \frac{2-y}{xy^2 - y + 1}.$$

Proof Consider the assumed generating function

$$O(y;x) = F_0(x) + F_1(x)y + F_2(x)y^2 + F_3(x)y^3 + \cdots,$$

and define

$$\begin{aligned} S_1(y;x) &= F_1(x) + F_2(x)y + F_3(x)y^2 + F_4(x)y^3 + \cdots, \\ S_2(y;x) &= F_2(x) + F_3(x)y + F_4(x)y^2 + F_5(x)y^3 + \cdots, \end{aligned}$$

which, using the known values for $F_0(x)$ and $F_1(x)$, can be written in terms of $O(y;x)$ as

$$\begin{aligned} S_1(y;x) &= \frac{1}{y}[O(y;x) - F_0(x)] = \frac{1}{y}[O(y;x) - 2], \\ S_2(y;x) &= \frac{1}{y^2}[O(y;x) - F_0(x) - F_1(x)y] = \frac{1}{y^2}[O(y;x) - 2 - y]. \end{aligned}$$

Substituting these into the functional equation

$$S_2(y;x) - S_1(y;x) + xO(y;x) = 0$$

consistent with the Lemma 1 recursion gives the result when solved for $O(y;x)$.□

**Corollary 1** By construction (this is a straightforward matter to check), $O(y;x)$ satisfies the P.D.E.

$$(2-y)(xy^2 - y + 1)\frac{\partial O(y;x)}{\partial y} - (xy^2 - 4xy + 1)O(y;x) = 0.$$

[1] In fact the sequence $\{F_0(-1), F_1(-1), F_2(-1), \ldots\}$ describes the sequence of Lucas numbers because of the initial values $F_0(-1) = 2$, $F_1(-1) = 1$. Moreover, the $(n+1)$th Lucas polynomial $L_n(x)$, say, is expressible as $L_n(x) = x^n F_n(-x^{-2})$, and its (known) generating function readily recovers Lemma 2 (reader exercise); $L_n(x)$ is defined in (A11) of the Appendix.

**Lemma 3** $F_n(x)$ has the closed form

$$F_n(x) = \frac{1}{2^n}\left[(1+\sqrt{1-4x})^n + (1-\sqrt{1-4x})^n\right], \qquad n \geq 0.$$

Proof A The characteristic equation for the recursion of Lemma 1 is $0 = \lambda^2 - \lambda + x$, with roots $\lambda_{1,2}(x) = \frac{1}{2}(1 \pm \sqrt{1-4x})$. For $\lambda_1 \neq \lambda_2$, the general solution

$$F_n(x) = A(\lambda_1)^n + B(\lambda_2)^n$$

for $n \geq 0$ yields the given form $F_n(x) = (\lambda_1)^n + (\lambda_2)^n$ as a particular solution upon applying the initial values for $F_0(x), F_1(x)$ ($\Rightarrow A = B = 1$).$\Box$

Proof B By definition, $x = pq = p(1-p) \Rightarrow p(x) = \frac{1}{2}(1 \pm \sqrt{1-4x})$, $q(x) = 1 - p(x) = \frac{1}{2}(1 \mp \sqrt{1-4x})$, from which Lemma 3 follows by (3).$\Box$

Lemma 3 Example: $n = 4$

$$\begin{aligned}
2^4 F_4(x) &= (1+\sqrt{1-4x})^4 + (1-\sqrt{1-4x})^4 \\
&= \left[1 + 4\sqrt{1-4x} + 6(1-4x) + 4(\sqrt{1-4x})^3 + (1-4x)^2\right] \\
&\quad + \left[1 - 4\sqrt{1-4x} + 6(1-4x) - 4(\sqrt{1-4x})^3 + (1-4x)^2\right] \\
&= 2[1 + 6(1-4x) + (1-4x)^2] \\
&= 16(1 - 4x + 2x^2) \\
\Rightarrow F_4(x) &= 1 - 4x + 2x^2. \qquad (5)
\end{aligned}$$

Remark 2 The excluded case $\lambda_1 = \lambda_2$ in Proof A corresponds to $x$ taking (maximum) value $\frac{1}{4}$. It gives a general solution $F_n(\frac{1}{4}) = (Cn+D)(\frac{1}{2})^n$, and in turn a particular one $F_n(\frac{1}{4}) = 2^{1-n}$ $(n \geq 0)$ in agreement with Lemma 3.

The following result has been obtained using the Maple package "FPS.mpl" (see [2,3] for more information) in combination with "hsum6.mpl" of the author W.A.K.[2] To find the assumed hypergeometric power series representation $F_n(x) = \sum_k A_k(n,x)x^k$, FPS.mpl first finds a linear differential equation for $F_n(x)$ which is converted into a linear recurrence equation for the coefficients $A_k(n,x)$; this is found to be of first order and hence $F_n(x)$ is a hypergeometric function. Each of Lemmas 1-3 given so far can also be generated by algebraic computation without difficulty.

**Lemma 4** For $n \geq 1$, $F_n(x)$ can be written as the explicit polynomial

$$F_n(x) = n \sum_{k=0}^{[\frac{1}{2}n]} \frac{1}{n-k} \binom{n-k}{k} (-x)^k,$$

[2]Available at **http://www.mathematik.uni-kassel.de/~koepf/Publikationen**.

the upper limit $[\frac{1}{2}n]$ denoting the greatest integer not exceeding $\frac{1}{2}n$.

Lemma 4 Example: $n = 6$

$$\begin{aligned} F_6(x) &= 6\sum_{k=0}^{3} \frac{1}{6-k} \binom{6-k}{k} (-x)^k \\ &= \binom{6}{0} (-x)^0 + \frac{6}{5} \binom{5}{1} (-x)^1 \\ &\qquad + \frac{3}{2} \binom{4}{2} (-x)^2 + 2 \binom{3}{3} (-x)^3 \\ &= 1 - 6x + 9x^2 - 2x^3. \end{aligned} \tag{6}$$

**Corollary 2** Associated with the Lemma 4 form of $F_n(x)$ is the hypergeometric representation (in standard notation)

$$F_n(x) = \lim_{\varepsilon \to 0} \left\{ {}_2F_1 \left( \begin{matrix} -\frac{1}{2}n, -\frac{1}{2}(n-1) \\ -(n-1)+\varepsilon \end{matrix} \,\middle|\, 4x \right) \right\}, \qquad n \geq 1.$$

Remark 3 Without the Corollary 2 interpretation of $F_n(x)$ as a limit, the hypergeometric series therein would otherwise be undefined strictly speaking because of terms of indeterminate form $\frac{0}{0}$.

Remark 4 Equating $F_n(x)$ as stated in Lemmas 3,4 gives an identity which holds for $n \geq 1$ and is a special case of Identity No. 1.64 in Gould's well known listing [4, p.8] with $z = -4x$. On a historical point, note that Identity No. 1.64 is itself a special case of equation (1) in Gould's 1999 paper "The Girard-Waring Power Sum Formulas for Symmetric Functions and Fibonacci Sequences" (*Fib. Quart.*, **37**(2), pp.135-140). The latter is but one of three identities (termed the Girard-Waring formulas) which, according to Gould, are particular instances of an even more general formula first found by Albert Girard in the 17th century and given later, in the 18th century, by Edward Waring for sums of powers of roots of a polynomial.

With reference to Remark 1, an interesting result we prove by induction is an analogue of the recurrence $F^*_{n+1}(x)F^*_{n-1}(x) - F^{*2}_n(x) = (-1)^n$ which exists for the aforementioned Fibonacci polynomials.

**Lemma 5** For $n \geq 1$,

$$F_{n+1}(x)F_{n-1}(x) - F_n^2(x) = (1-4x)x^{n-1}.$$

Proof A The result holds for $n = 1$, with each side of the equation $1 - 4x$. Suppose it is true for $n = k \geq 1$, and consider the case when $n = k + 1$.

Then

$$\begin{aligned} F_{k+2}(x)F_k(x) &= [F_{k+1}(x) - xF_k(x)]F_k(x) \\ &= F_{k+1}(x)[F_{k+1}(x) + xF_{k-1}(x)] - xF_k^2(x) \end{aligned}$$

using Lemma 1 twice (*i.e.*, for $n = k, k-1$ respectively). Continuing,

$$\begin{aligned} &= F_{k+1}^2(x) + [F_{k+1}(x)F_{k-1}(x) - F_k^2(x)]x \\ &= F_{k+1}^2(x) + (1-4x)x^k \end{aligned}$$

by assumption, and so the inductive step is upheld.□

Proof B For $x \neq \frac{1}{4}$ then, from Proof A of Lemma 3, $F_n(x) = (\lambda_1)^n + (\lambda_2)^n$ with $\lambda_1 + \lambda_2 = 1$ and $\lambda_1\lambda_2 = x$ ($\lambda_1 \neq \lambda_2$), whence

$$\begin{aligned} &F_{n+2}(x)F_n(x) - F_{n+1}^2(x) \\ &\quad = [(\lambda_1)^{n+2} + (\lambda_2)^{n+2}][(\lambda_1)^n + (\lambda_2)^n] - [(\lambda_1)^{n+1} + (\lambda_2)^{n+1}]^2 \\ &\quad = (\lambda_1)^n(\lambda_2)^n[(\lambda_1)^2 + (\lambda_2)^2 - 2\lambda_1\lambda_2] \\ &\quad = (\lambda_1\lambda_2)^n[(\lambda_1+\lambda_2)^2 - 4\lambda_1\lambda_2] \\ &\quad = x^n(1-4x).\square \end{aligned}$$

Lemma 5 Example: $n = 4$

$$\begin{aligned} \text{l.h.s.} &= F_5(x)F_3(x) - F_4^2(x) \\ &= (1-5x+5x^2)(1-3x) - (1-4x+2x^2)^2 \\ &= 1-8x+20x^2-15x^3 - (1-8x+20x^2-16x^3+4x^4) \\ &= x^3(1-4x) \\ &= \text{r.h.s.} \end{aligned} \tag{7}$$

## Particular Sequences

Here we examine—purely from a mathematical point of interest—three sequences $\{F_0(x), F_1(x), F_2(x), \ldots\}$ arising from particular values of $x$.

I: $x = \frac{1}{4}$

The sequence $\{F_0(\frac{1}{4}), F_1(\frac{1}{4}), F_2(\frac{1}{4}), \ldots\}$, with general term $F_n(\frac{1}{4}) = 2^{1-n}$ for $n \geq 0$ (see Remark 2), satisfies the non-linear Lemma 5 recursion trivially, as it does that of Lemma 1. We also note that

$$\lim_{n\to\infty} \left\{ \frac{F_n(\frac{1}{4})}{F_{n-1}(\frac{1}{4})} \right\} = \frac{1}{2}, \tag{8}$$

the ratio $F_n(\frac{1}{4})/F_{n-1}(\frac{1}{4})$ itself taking this constant value for $n \geq 1$.

<u>II: $x = 1$</u>

The sequence $\{F_0(1), F_1(1), F_2(1), \ldots\}$ has, by Lemma 3, general term

$$\begin{aligned} F_n(1) &= \frac{1}{2^n}\left[(1+\sqrt{3}i)^n + (1-\sqrt{3}i)^n\right] \\ &= \begin{cases} 2(-1)^n & 3 \mid n \\ (-1)^{n-1} & \text{otherwise} \end{cases} \end{aligned} \tag{9}$$

for $n \geq 0$ (see also, in the light of Lemma 4, Identity No. 1.68 of [4]). The closed form here is easy to establish once it is realised that $\frac{1}{2}(1 \pm \sqrt{3}i)$ are the two complex cube roots of $-1$, since it is then immediate that $F_{n+3}(1) = -F_n(1)$. Using this for values of $n \geq 0$ in conjunction with starting values $F_0(1) = 2$, $F_1(1) = 1$ and $F_2(1) = -1$, the result follows easily by inspection. In this case the ratio of terms $F_n(1)/F_{n-1}(1)$ has no limit for large $n$, for it takes the self-repeating values $\frac{1}{2}, -1, 2, \frac{1}{2}, -1, 2, \ldots,$ as $n = 1, 2, 3, \ldots$

<u>Remark 5</u> Since we have a closed form, as a further check for this case we show that the general term (9) indeed satisfies the linear recursion derived at the start of the section. Suppose, firstly, $3 \mid n$, so that $F_n(1) = 2(-1)^n$ with (3 thus being a factor of neither $n+1$ nor $n+2$) $F_{n+1}(1) = (-1)^{(n+1)-1} = (-1)^n$ and $F_{n+2}(1) = (-1)^{(n+2)-1} = (-1)^{n+1}$, whence Lemma 1 holds. If, on the other hand, 3 does not divide $n$ then $F_n(1) = (-1)^{n-1}$ and there are two further cases to consider: (i) $3 \mid n+2$ but not $n+1$ (in which case $F_{n+1}(1) = (-1)^n$ and $F_{n+2}(1) = 2(-1)^n$), or (ii) $3 \mid n+1$ but not $n+2$ (giving $F_{n+1}(1) = 2(-1)^{n+1}$, $F_{n+2}(1) = (-1)^{n+1}$); both render Lemma 1 true. Lemma 5 can be validated by a similar type of argument which we leave as an exercise for the keen reader.

<u>III: $x = -1$</u>

We state the following merely for completeness. Lemma 3 gives

$$F_n(-1) = \frac{1}{2^n}\left[(1+\sqrt{5})^n + (1-\sqrt{5})^n\right], \qquad n \geq 0, \tag{10}$$

as the general term of the sequence $\{F_0(-1), F_1(-1), F_2(-1), \ldots\}$ of Lucas numbers. Writing $\phi = \frac{1}{2}(1+\sqrt{5})$, $\hat{\phi} = \frac{1}{2}(1-\sqrt{5})$, then $F_n(-1) = \phi^n + \hat{\phi}^n$ and, noting that $|\hat{\phi}/\phi| < 1$,

$$\lim_{n\to\infty}\left\{\frac{F_n(-1)}{F_{n-1}(-1)}\right\} = \lim_{n\to\infty}\left\{\frac{\phi^n + \hat{\phi}^n}{\phi^{n-1} + \hat{\phi}^{n-1}}\right\}$$

$$
\begin{aligned}
&= \phi \lim_{n\to\infty} \left\{ \frac{1+(\hat{\phi}/\phi)^n}{1+(\hat{\phi}/\phi)^{n-1}} \right\} \\
&= \phi, \qquad\qquad (11)
\end{aligned}
$$

the celebrated "golden ratio" associated with the Fibonacci and Lucas numbers; this is to be expected in view of Remark 1.

Rather than extend this section any further, we direct the reader to the Appendix where other characteristics of the function $F_n(x)$ are given. It should be noted that such a function is actually a particular case of one of the so called Dickson polynomials of considerable historical standing.

# A New Catalan Number Representation

We finish with a short section which details our main result of the paper, as indicated in the Introduction. Employing (3) we write $E_n$ (2) as

$$
E_n = n \sum_{k=0}^{n-1} \binom{2n-k-1}{n-k-1} x^{n-k-1} F_{k+1}(x), \qquad n \geq 1, \qquad (12)
$$

after a simple shift in the summing index. Lemma 4 allows us to identify the coefficient of $x^\alpha$ in $E_n$ which, when compared to that of the Theorem of Shapiro and Hamilton, yields the following representation of the general Catalan number (it can also be read, of course, as a binomial coefficient identity in the sense that the l.h.s. is evaluated). For $n \geq 1$, $0 \leq \alpha \leq n-1$,

$$
\begin{aligned}
&\frac{(-1)^{\alpha+1-n}}{n-\alpha} \sum_{k=n-\alpha-1}^{n-1} (-1)^k (k+1) \binom{n-\alpha}{k+\alpha+1-n} \binom{2n-k-1}{n-k-1} \\
&= c_\alpha; \qquad\qquad (13)
\end{aligned}
$$

the lower limit on the sum manifests itself naturally as $k = 0$, but has been modified to that shown so as to eliminate non-contributing terms in the sum arising from the first binomial coefficient. This result is unusual at first sight, and it is appropriate to provide an example by way of illustration:

<u>Example: $n = 7$, $\alpha = 3$</u>

$$
\begin{aligned}
\text{l.h.s.} &= \frac{(-1)^{-3}}{4} \sum_{k=3}^{6} (-1)^k (k+1) \binom{4}{k-3} \binom{13-k}{6-k} \\
&= -\frac{1}{4} \left[ -4 \binom{4}{0} \binom{10}{3} + 5 \binom{4}{1} \binom{9}{2} \right.
\end{aligned}
$$

$$
\begin{aligned}
& \qquad - 6\binom{4}{2}\binom{8}{1} + 7\binom{4}{3}\binom{7}{0}\Bigg] \\
= & \; -\frac{1}{4}[-4\cdot 1\cdot 120 + 5\cdot 4\cdot 36 - 6\cdot 6\cdot 8 + 7\cdot 4\cdot 1] \\
= & \; -\frac{1}{4}[-480 + 720 - 288 + 28] \\
= & \; 5 \\
= & \; c_3 \\
= & \; \text{r.h.s.} \qquad (14)
\end{aligned}
$$

The identity lends itself readily to simplification to a final form. It is non-standard in so far as what looks to be a function of two variables is actually univariate, although it is by no means unique in kind—such results do crop up from time to time in combinatorics.

**Theorem** For $n \geq 1$, $0 \leq \alpha \leq n-1$,

$$
c_\alpha = \frac{1}{n-\alpha}\sum_{k=0}^{\alpha}(-1)^k(k+n-\alpha)\binom{n-\alpha}{k}\binom{n+\alpha-k}{n}.
$$

Remark 6 An independent analytic proof of the Theorem is unnecessary, but we have validated it by computation in fully symbolic form. The package hsum6.mpl returns a hypergeometric representation

$$
\binom{n+\alpha}{n}\, {}_3F_2\left(\begin{matrix} n-\alpha+1, -\alpha, -(n-\alpha) \\ n-\alpha, -(n+\alpha) \end{matrix}\,\Bigg|\, 1\right) \tag{15}
$$

for its complete r.h.s., and duly reduces it to the required l.h.s. using Zeilberger's algorithm (see [5]).

Remark 7 If $n = 1$ then $\alpha = 0$ and the r.h.s. of the Theorem contracts trivially to the single term $\binom{1}{0}\binom{1}{1} = 1 = c_0$, which is correct. For $n \geq 2$, the special case when $\alpha$ assumes its highest value $n-1$ ($\Rightarrow \alpha \geq 1$) gives an interesting form of the Catalan number $c_\alpha$ as a difference in binomial coefficients; we have that

$$
\begin{aligned}
c_\alpha &= \sum_{k=0}^{\alpha}(-1)^k(k+1)\binom{1}{k}\binom{2\alpha+1-k}{\alpha+1} \\
&= \binom{2\alpha+1}{\alpha+1} - 2\binom{2\alpha}{\alpha+1}, \qquad (16)
\end{aligned}
$$

which is not often seen (and actually holds for $\alpha \geq 0$).

# Summary

A previous result established elsewhere in the context of a theoretical World Series type event has led to a new representation of the general Catalan number;[3] note that an algebraic proof of our Theorem, together with an inverse formula (and also generalisations of these to accommodate so called "ballot numbers" $a_{nm} = \frac{n+1-m}{n+1}\binom{n+m}{m}$ of which $c_n = a_{nn}$ is a special case), are offered by Gould in this volume of *Congressus Numerantium*—see the accompanying article "Proof and Generalization of a Catalan Number Formula of Larcombe". On a wider point, it remains to be seen whether or not an enhanced version of the sort of competition described here gives any further results involving Catalan numbers. Certainly, in terms of (as considered here) the expected number of trials before declaration of a winner, the two special cases of Theorem 4 in Siegrist's treatment of an $n$-point, win-by-$k$ game (see (8),(9) on p.812 of [9]) suggests this as a potential topic for future study.

# Appendix

Here we detail additional properties of the function $F_n(x)$ which is also linked, for completeness, to a Dickson polynomial of which it describes a special case.

Consider the definition of $F_n(x)$ (3). Writing instead (since $x = pq$, $p+q = 1$)

$$F_n(p - p^2) = p^n + (1-p)^n \tag{A1}$$

and differentiating twice w.r.t. $p$ gives

$$(1-4x)F_n''(x) - 2F_n'(x) = n(n-1)F_{n-2}(x), \qquad n \geq 2. \tag{A2}$$

Differentiating twice more, then twice more again, leads to the further relations

$$\begin{aligned}(1-4x)^2F_n''''(x) - 12(1-4x)F_n'''(x) + 12F_n''(x)\\ = n(n-1)(n-2)(n-3)F_{n-4}(x)\end{aligned} \tag{A3}$$

[3]It is perhaps worth mentioning that the Theorem of Shapiro and Hamilton was formulated, in fact, in a paper by Maisel as early as 1966 (see (3.4) on p.330 of [6]); this may well be its first appearance in the literature. Note also that their expression $E_n(x) = n\sum_{k=0}^{n-1} c_k x^k$ is given a closed form in Lengyel [7, Theorem 1, p.295] for $x = \frac{1}{4}$ (it would seem, from computer experimentation, that no other values of $x \in (0, \frac{1}{4}]$ admit the sum to be evaluated in such a way). We remark too that Draim and Bicknell, in an article [8] also of 1966, considered roots of a "primitive" equation which is slightly more general than that associated with the Lemma 1 recurrence $\lambda^2 - \lambda + x = 0$.

and

$$\begin{aligned}(1-4x)^3 F_n''''''(x) - 30(1-4x)^2 F_n'''''(x) \\ + 180(1-4x)F_n''''(x) - 120F_n'''(x) \\ = n(n-1)(n-2)(n-3)(n-4)(n-5)F_{n-6}(x). \quad \text{(A4)}\end{aligned}$$

After taking the $(2r)$th derivative of (A1) $(r \geq 1)$ the resulting differential equation will, for $n \geq 2r$, assume a form

$$\sum_{i=r}^{2r} \alpha_i^{(r)} (1-4x)^{i-r} F_n^{[i]}(x) = \left[\prod_{i=0}^{2r-1} (n-i)\right] F_{n-2r}(x), \qquad \text{(A5)}$$

where $F_n^{[i]}(x)$ denotes the $i$th derivative of $F_n(x)$ w.r.t. $x$ and $\alpha_i^{(r)}$ is a constant. Note that $\alpha_{2r}^{(r)} = 1$ (by inspection when (A2)-(A4) have been worked through).

A different type of result may be formulated by instead integrating (A1) w.r.t. $p$ twice, yielding (for $\beta$ constant)

$$F_{n+2}(x) = (n+1)(n+2) \int \frac{g(x)}{\sqrt{1-4x}}\, dx - \beta \left[1 - \sqrt{1-4x}\right], \qquad \text{(A6)}$$

where

$$g(x) = \int \frac{F_n(x)}{\sqrt{1-4x}}\, dx. \qquad \text{(A7)}$$

As an example, $n = 0$ produces eventually from this scheme the polynomial $F_2(x) = C_1 - 2x + C_2\sqrt{1-4x}$ which is correct for constants $C_1 = 1$, $C_2 = 0$.

Other identities can be obtained routinely by taking powers of (3), the first few examples of which are, for $n \geq 0$,

$$\begin{aligned} F_n^2(x) &= F_{2n}(x) + 2x^n, \\ F_n^3(x) &= F_{3n}(x) + 3x^n F_n(x), \\ F_n^4(x) &= F_{4n}(x) + 4x^n F_{2n}(x) + 6x^{2n}, \\ F_n^5(x) &= F_{5n}(x) + 5x^n F_{3n}(x) + 10x^{2n} F_n(x), \qquad \text{(A8)} \end{aligned}$$

*etc*; a general representation of $F_n^s(x)$ is readily available for $s \geq 2$ (left as a routine reader exercise).

To end, we remark that the polynomial $F_n(x)$ is a particular instance of what is known as a Dickson polynomial of the first kind (of degree $n$ in $x$, with real parameter $a$) defined as

$$D_n(x,a) = n \sum_{k=0}^{[\frac{1}{2}n]} \frac{1}{n-k} \left( \begin{array}{c} n-k \\ k \end{array} \right) (-a)^k x^{n-2k}; \tag{A9}$$

clearly

$$F_n(x) = D_n(1,x), \qquad n \geq 0, \tag{A10}$$

with the (standard) initial values $D_0(x,a) = 2$, $D_1(x,a) = x$ corresponding to $F_0(x) = 2$, $F_1(x) = 1$. As a point of interest, it is easily shown that, for indeterminates $u_1, u_2$, $D_n(u_1 + u_2, u_1 u_2) = (u_1)^n + (u_2)^n$ which reads as (A10) for $u_1 = p$, $u_2 = q$. Note also that setting $a = -1$ in (A9) gives

$$D_n(x,-1) = n \sum_{k=0}^{[\frac{1}{2}n]} \frac{1}{n-k} \left( \begin{array}{c} n-k \\ k \end{array} \right) x^{n-2k} = L_n(x), \tag{A11}$$

which is the general Lucas polynomial mentioned briefly in Footnote 1 and from which we may write $D_n(x,-1) = x^n F_n(-x^{-2})$—in other words, as an alternative to (A10),

$$F_n(x) = (-x)^{\frac{1}{2}n} D_n((-x)^{-\frac{1}{2}}, -1), \qquad n \geq 0, \tag{A12}$$

equality at $n = 0, 1$ guaranteed by the definition of $D_0(x,a), D_1(x,a)$.

Dickson polynomials date from the Ph.D. study of L.E. Dickson at the close of the 19th century. Algebraic, arithmetic and analytic properties of Dickson polynomials of the first and second kind (and their generalisations to several indeterminates), along with applications, have been drawn together in an authoritative monograph by Lidl *et al.* [10] to which we refer now having noted (A10)—in particular, (i) the closed form for $D_n(x,a)$ [10, p.9] contracts immediately to ours for $F_n(x)$ in Lemma 3, (ii) Lemma 1 is likewise available from the recurrence for $D_n(x,a)$ [10, Lemma 2.3, p.10], and (iii) the ordinary generating function for $D_n(x,a)$ [10, Lemma 2.4, p.10] recovers that for $F_n(x)$ stated here as Lemma 2; it should be pointed out that the methods we have used to establish Lemmas 1-3 differ sufficiently to those presented by Lidl *et al.* in relation to $D_n(x,a)$, and so justify their inclusion.

# References


[1] Shapiro, L. and Hamilton, W. (1993). The Catalan numbers visit the world series, *Math. Mag.*, **66**, pp.20-22.

[2] Koepf, W. (1992). Power series in computer algebra, *J. Sym. Comp.*, **13**, pp.581-603.

[3] Gruntz, D. and Koepf, W. (1995). Maple package on formal power series, *Maple Tech. News.*, **2**(2), pp.22-28.

[4] Gould, H.W. (1972). Combinatorial identities, Rev. Ed., University of West Virginia, U.S.A.

[5] Koepf, W. (1998). Hypergeometric summation: an algorithmic approach to summation and special function identities, Vieweg, Wiesbaden, Germany.

[6] Maisel, H. (1966). Best $k$ of $2k-1$ comparisons, *J. Amer. Stat. Ass.*, **61**, pp.329-344.

[7] Lengyel, T. (1993). A combinatorial identity and the world series, *S.I.A.M. Rev.*, **35**, pp.294-297.

[8] Draim, N.A. and Bicknell, M. (1966). Sums of $n$-th powers of roots of a given quadratic equation, *Fib. Quart.*, **4**(2), pp.170-178.

[9] Siegrist, K. (1989). $n$-point, win-by-$k$ games, *J. Appl. Prob.*, **27**, pp.807-814.

[10] Lidl, R., Mullen, G.L. and Turnwald, G. (1993). Dickson polynomials (Pitman Monographs and Surveys in Pure and Applied Mathematics, No. 65), Longman, London, U.K.