The Curious History of Faà di Bruno's Formula

Warren P. Johnson

1. WHAT IS THE *m*th DERIVATIVE OF A COMPOSITE FUNCTION? By far the best known answer is

Faà di Bruno's Formula. *If g and f are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m} g(f(t)) =$$

$$\sum \frac{m!}{b_1!\, b_2! \cdots b_m!} g^{(k)}(f(t)) \left(\frac{f'(t)}{1!}\right)^{b_1} \left(\frac{f''(t)}{2!}\right)^{b_2} \cdots \left(\frac{f^{(m)}(t)}{m!}\right)^{b_m}, \quad (1.1)$$

where the sum is over all different solutions in nonnegative integers $b_1, \ldots, b_m$ of $b_1 + 2b_2 + \cdots + mb_m = m$, and $k := b_1 + \cdots + b_m$.

For example, when $m = 3$ this instructs us to look at all solutions in nonnegative integers of the equation $b_1 + 2b_2 + 3b_3 = 3$. We can have $b_3 = 1$, $b_1 = 0 = b_2$, in which case $k = 1$ and we get the term

$$\frac{3!}{0!\,0!\,1!} g'(f(t)) \left(\frac{f'''(t)}{3!}\right) = g'(f(t))\, f'''(t).$$

Otherwise we must have $b_3 = 0$. Then we can have $b_2 = 1 = b_1$, in which case $k = 2$ and we get the term

$$\frac{3!}{1!\,1!\,0!} g''(f(t)) \left(\frac{f'(t)}{1!}\right) \left(\frac{f''(t)}{2!}\right) = 3g''(f(t))\, f'(t)\, f''(t).$$

The only other solution is $b_1 = 3$, $b_2 = 0 = b_3$, in which case $k = 3$ and we get the term

$$\frac{3!}{3!\,0!\,0!} g'''(f(t)) \left(\frac{f'(t)}{1!}\right)^3 = g'''(f(t)) \left(f'(t)\right)^3.$$

Therefore Faà di Bruno's formula says (correctly) that

$$\frac{d^3}{dt^3} g(f(t)) = g'(f(t))\, f'''(t) + 3g''(f(t))\, f'(t)\, f''(t) + g'''(f(t)) \left(f'(t)\right)^3. \quad (1.2)$$

In spite of its appearance, (1.1) is rather simple when conceived of in the right way, as was recently pointed out in this MONTHLY by Harley Flanders [**24**]. A restatement in terms of set partitions can be proved easily in a few lines, as we shall see in Section 2, though it still requires a bit of work to pass from that form to the form in (1.1).

Once Faà di Bruno's formula was considered a real analysis result: it is in the *Cours d'Analyse* of Goursat [**28**] and of de la Vallée Poussin [**66**]. Riordan and Comtet [**53**], [**12**], [**13**] saw it as part of combinatorial analysis, a term that seems to be going out of fashion; the subject subsumed in algebraic combinatorics, the books of Riordan and Comtet largely superseded by Stanley's monumental [**64**] and [**65**], where Faà di Bruno's formula is mentioned [**65**, p. 65], but not stated. It can also be found in books on partitions [**3**], mathematical statistics [**14**], matrix theory [**35**], calculus of finite differences [**37**], computer science [**38**], symmetric functions [**44**], and miscellaneous mathematical techniques [**46**].

Faà di Bruno published his formula in [**16**] and [**17**], which both date from December 1855. [**16**] is in Italian and [**17**] in French, but otherwise they are essentially the same—both are very short, containing little more than the statement of the result. The usual reference for his demonstration is the appendix of his best known book [**20**], which would not appear for another 20 years. The same proof (an induction that has left some commentators unsatisfied) is in the appendix of [**19**], which came out in 1859.

But there is much more to the story. The "little more" in [**16**] and [**17**] includes a determinantal version of Faà di Bruno's formula, which has received little attention. That several other mathematicians found different expressions for the mth derivative of $g(f(t))$ in the 19th century has been forgotten; these are all independent of Faà di Bruno's work and a few of them predate it. Most of all, Faà di Bruno was neither the first to state the formula that bears his name, nor the first to prove it. We elaborate on all this below.

2. A COMBINATORIAL ARGUMENT. It is convenient to begin by discussing the *Bell polynomials*, which are associated with set partitions. To the partition $\{1\}$ we associate the monomial x_1; this is the only partition of the set $\{1\}$, and we define $\mathbf{B}_{1,1}(x_1) = x_1$. The set $\{1, 2\}$ has the two partitions $\{1, 2\}$ and $\{1\}, \{2\}$, the former with one block and the latter with two, and we associate to them the monomials x_2 and x_1^2, respectively. Then $\mathbf{B}_{2,1}(x_1, x_2) = x_2$ and $\mathbf{B}_{2,2}(x_1) = x_1^2$.

There are five partitions of the set $\{1, 2, 3\}$. Three of these have two blocks, namely $\{1, 2\}, \{3\}$ and $\{1, 3\}, \{2\}$ and $\{1\}, \{2, 3\}$; we associate the monomial x_1x_2 to each of these, and so $\mathbf{B}_{3,2}(x_1, x_2) = 3x_1x_2$. The other Bell polynomials of order three are $\mathbf{B}_{3,3}(x_1) = x_1^3$, corresponding to $\{1\}, \{2\}, \{3\}$; and $\mathbf{B}_{3,1}(x_1, x_2, x_3) = x_3$, corresponding to $\{1, 2, 3\}$. In general,

$$\mathbf{B}_{m,k}(x_1, x_2, \ldots, x_{m-k+1}) = \frac{1}{k!} \sum_{\substack{j_1+\cdots+j_k=m \\ j_i\geq 1}} \binom{m}{j_1, \ldots, j_k} x_{j_1} \cdots x_{j_k},$$

where $\mathbf{B}_{0,0}(x_1) = 1$; the sum is effectively over set partitions of $\{1, 2, \ldots, m\}$ with block sizes $j_1, \ldots, j_k$, with the factor $1/k!$ correcting for the multiple counting inside the sum. Only $m - k + 1$ variables are necessary because no block can contain more than $m - k + 1$ elements. A further example is $\mathbf{B}_{4,2}(x_1, x_2, x_3) = 4x_1x_3 + 3x_2^2$, where there are 7 partitions of $\{1, 2, 3, 4\}$ into two blocks, 4 with one block of size three and one of size one, and 3 with two blocks of size two.

We pause for a remark to be used later. If we set every $x_i = 1$, we are simply counting the number of partitions of $\{1, 2, \ldots, m\}$ into k blocks, which is the Stirling number of the second kind. Following [**39**] (which anyone interested in Stirling numbers should consult) I denote these by $\left\{ {m \atop k} \right\}$; thus

$$\mathbf{B}_{m,k}(1, 1, \ldots, 1) = \left\{ {m \atop k} \right\}. \tag{2.1}$$

This approach to the Bell polynomials seems to have originated in [**26**], which was published in a Chilean journal by Roberto Frucht and Gian-Carlo Rota in 1965. (Frucht's later paper [**25**] may be more accessible to some readers.) The name "Bell polynomials" was introduced by Riordan in [**53**], and it was he who first observed that they are ideally suited to the description of Faà di Bruno's formula [**52**]. The polynomials that Bell actually considered in [**5**] were

$$\mathbf{Y}_m = \mathbf{Y}_m(x_1, x_2, \ldots, x_m) := \sum_{k=0}^{m} \mathbf{B}_{m,k}(x_1, \ldots, x_{m-k+1}).$$

Thus, for example, $\mathbf{Y}_3(x_1, x_2, x_3) = x_3 + 3x_1x_2 + x_1^3$. Note that one can recover the $\mathbf{B}_{m,k}$'s from the $\mathbf{Y}_m$'s by grouping the terms of $\mathbf{Y}_m$ according to degree.

Now associate set partitions to derivatives of composite functions in the same way. To $g'(f(t))\,f'(t)$ we associate $\{1\}$. The partitions $\{1, 2\}$ and $\{1\}, \{2\}$ correspond to $g'(f(t))\,f''(t)$ and $g''(f(t))(f'(t))^2$, respectively. To each partition of $\{1, 2, \ldots, m\}$ with k blocks corresponds a term of $d^m\, g(f(t))/dt^m$ with the factor $g^{(k)}(f(t))$, where the block sizes determine its other factors (which are derivatives of f). Thus the partitions $\{1, 2\}, \{3\}$ and $\{1, 3\}, \{2\}$ and $\{1\}, \{2, 3\}$, each with one block of size 1 and one of size 2, each correspond to $g''(f(t))\,f'(t)\,f''(t)$, and the other two partitions of $\{1, 2, 3\}$, namely $\{1, 2, 3\}$ itself and $\{1\}, \{2\}, \{3\}$, are associated to $g'(f(t))\,f'''(t)$ and $g'''(f(t))(f'(t))^3$ respectively; adding these five terms we get the right side of (1.2). The general result is

Faà di Bruno's formula, set partition version. *If g and f are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m}\, g\,(f(t)) = \sum g^{(k)}\,(f(t))\,\left(f'(t)\right)^{b_1}\,\left(f''(t)\right)^{b_2}\cdots\left(f^{(m)}(t)\right)^{b_m},$$

where the sum is over all partitions of $\{1, 2, \ldots, m\}$, *and, for each partition, k is its number of blocks and b_i is the number of blocks with exactly i elements.*

We may prove this by induction on m. Every partition of $\{1, 2, \ldots, m+1\}$ can be obtained in a unique way by adjoining $m+1$ to a partition of $\{1, 2, \ldots, m\}$. If we add $\{m+1\}$ as a new singleton block, then we increase the number of blocks of size 1 by one, and the total number of blocks by one. This corresponds to applying d/dt to $g^{(k)}(f(t))$ to get $g^{(k+1)}(f(t))\,f'(t)$. On the other hand, if we add $m+1$ to an existing block of size i (say), then the number of such blocks decreases by one, the number of blocks of size $i+1$ increases by one, and the total number of blocks remains the same. If we started with b_i blocks of size i, then we may add $m+1$ to any of them to produce this effect. This corresponds to applying d/dt to $(f^{(i)}(t))^{b_i}$ to get $b_i(f^{(i)}(t))^{b_i-1}f^{(i+1)}(t)$; hence the result. An immediate corollary is

Faà di Bruno's formula, Bell polynomial version (Riordan's formula). *If g and f are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m}\, g(f(t)) = \sum_{k=0}^{m} g^{(k)}\,(f(t))\,\mathbf{B}_{m,k}\left(f'(t), f''(t), \ldots, f^{(m-k+1)}(t)\right). \tag{2.2}$$

As such, this dates back to [**52**], but Cesàro gave a similar statement in terms of his *calcul isobarique* [**10**], so one might quibble with the name "Riordan's formula". (Faà di Bruno used the same word *isobariche*, which was suggested to him by Cayley, in describing his eponymous formula in [**18**], though if Cesàro knew this, he did not say so.) The symmetrized form of Faà di Bruno's formula implied by (2.2) dates back at least to Hess in 1872 [**31**]. The classical Faà di Bruno formula (1.1) differs only in combining like terms. The number of partitions of $\{1, 2, \ldots, m\}$ into b_1 1-blocks, b_2 2-blocks, etc. would be

$$\binom{m}{\underbrace{1,\ldots,1}_{b_1\ 1'\text{s}},\underbrace{2,\ldots,2}_{b_2\ 2'\text{s}},\underbrace{3,\ldots,3}_{b_3\ 3'\text{s}},\ldots}$$

except that this makes artificial distinctions among the i-blocks for each i. The actual number of such partitions is

$$\binom{m}{\underbrace{1,\ldots,1}_{b_1\ 1'\text{s}},\underbrace{2,\ldots,2}_{b_2\ 2'\text{s}},\underbrace{3,\ldots,3}_{b_3\ 3'\text{s}},\ldots}\frac{1}{b_1!\,b_2!\ldots b_m!}$$

and (1.1) follows.

I think that this proof of Faà di Bruno's formula (with or without the material on Bell polynomials) would fit very well in an undergraduate course in combinatorics. It is surprising that such a simple argument is not better known. Besides [**26**] and [**25**], one can find another form of it in [**11**], and a q-version in [**36**]. Zeilberger suggested a (superficially) different combinatorial approach in [**69**]. Most induction proofs, such as the one in [**19**] and [**20**], are described in terms of number-theoretic (rather than set-theoretic) partitions, which makes the induction less transparent.

We may also use set partitions to prove Faà di Bruno's determinantal form of the result.

Lemma. *If $m \geq 1$, then*

$$\mathbf{Y}_m(x_1, x_2, \ldots, x_m) = \begin{vmatrix} \binom{m-1}{0}x_1 & \binom{m-1}{1}x_2 & \binom{m-1}{2}x_3 & \cdots & \binom{m-1}{m-2}x_{m-1} & \binom{m-1}{m-1}x_m \\ -1 & \binom{m-2}{0}x_1 & \binom{m-2}{1}x_2 & \cdots & \binom{m-2}{m-3}x_{m-2} & \binom{m-2}{m-2}x_{m-1} \\ 0 & -1 & \binom{m-3}{0}x_1 & \cdots & \binom{m-3}{m-4}x_{m-3} & \binom{m-3}{m-3}x_{m-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{1}{0}x_1 & \binom{1}{1}x_2 \\ 0 & 0 & 0 & \cdots & -1 & \binom{0}{0}x_1 \end{vmatrix}.$$

All entries on the main subdiagonal are -1, and all entries below it are 0. For example, $\mathbf{Y}_1(x_1) = x_1$, and

$$\mathbf{Y}_2(x_1, x_2) = \begin{vmatrix} x_1 & x_2 \\ -1 & x_1 \end{vmatrix} = x_1^2 + x_2.$$

As before, the term x_1^2 corresponds to the partition $\{1\}, \{2\}$, and the term x_2 to the partition $\{1, 2\}$. From this point of view, it is interesting to expand these determinants across the top row:

$$\begin{aligned} \mathbf{Y}_3(x_1, x_2, x_3) &= \begin{vmatrix} x_1 & 2x_2 & x_3 \\ -1 & x_1 & x_2 \\ 0 & -1 & x_1 \end{vmatrix} \\ &= x_1 \begin{vmatrix} x_1 & x_2 \\ -1 & x_1 \end{vmatrix} - 2x_2 \begin{vmatrix} -1 & x_2 \\ 0 & x_1 \end{vmatrix} + x_3 \begin{vmatrix} -1 & x_1 \\ 0 & -1 \end{vmatrix} \\ &= x_1(x_1^2 + x_2) + 2x_2(x_1) + x_3(1). \end{aligned}$$

While we could simplify this further, in this form we can see that the top row represents the blocks that contain the largest element. The term $x_1(x_1^2 + x_2)$ represents all the partitions of $\{1, 2, 3\}$ containing $\{3\}$ as a singleton block; specifically, x_1^3 corresponds to $\{1\}, \{2\}, \{3\}$ and $x_1 x_2$ to $\{1, 2\}, \{3\}$. The term $2x_2(x_1)$ corresponds to the partitions in which 3 is in a block of size two, namely $\{1, 3\}, \{2\}$ and $\{1\}, \{2, 3\}$; and the term $x_3(1)$ corresponds to the partition $\{1, 2, 3\}$. If we treat similarly the determinant for $\mathbf{Y}_4(x_1, x_2, x_3, x_4)$, we reach

$$\mathbf{Y}_4(x_1, x_2, x_3, x_4) = x_1(x_1^3 + 3x_1x_2 + x_3) + 3x_2(x_1^2 + x_2) + 3x_3(x_1) + x_4(1).$$

The term $x_1(x_1^3 + 3x_1x_2 + x_3)$ corresponds to the five partitions of $\{1, 2, 3, 4\}$ containing $\{4\}$ as a singleton block, the term $3x_2(x_1^2 + x_2)$ to the six partitions where 4 is in a block of size two, the term $3x_3(x_1)$ to the three partitions where 4 is in a block of size three, and the term $x_4(1)$ to the partition $\{1, 2, 3, 4\}$.

In general, if we expand the determinant across the top row we will get a sum $\sum_{k=1}^{m} \binom{m-1}{k-1} x_k\, c_{m,k}$, where $c_{m,k}$ is the cofactor of $\binom{m-1}{k-1}x_k$; thus, for example, $c_{4,2} = x_1^2 + x_2$. Note that $\binom{m-1}{k-1}$ is the number of ways we can choose the $k-1$ elements that are in the same block as the largest element m. It follows that $c_{m,k}$ represents the partitions of the remaining $m-k$ elements.

We use this combinatorial interpretation to prove the lemma by induction on m. We have already verified it up to $m = 4$. Assuming it holds for m, we have to show that

$$\mathbf{Y}_{m+1}(x_1, x_2, \ldots, x_{m+1}) \\ = \begin{vmatrix} \binom{m}{0}x_1 & \binom{m}{1}x_2 & \binom{m}{2}x_3 & \cdots & \binom{m}{m-1}x_m & \binom{m}{m}x_{m+1} \\ -1 & \binom{m-1}{0}x_1 & \binom{m-1}{1}x_2 & \cdots & \binom{m-1}{m-2}x_{m-1} & \binom{m-1}{m-1}x_m \\ 0 & -1 & \binom{m-2}{0}x_1 & \cdots & \binom{m-2}{m-3}x_{m-2} & \binom{m-2}{m-2}x_{m-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{1}{0}x_1 & \binom{1}{1}x_2 \\ 0 & 0 & 0 & \cdots & -1 & \binom{0}{0}x_1 \end{vmatrix}.$$

Expanding down the first column, we get

$$x_1\,\mathbf{Y}_m(x_1,\ldots,x_m)+\left|\begin{matrix}\binom{m}{1}x_2 & \binom{m}{2}x_3 & \cdots & \binom{m}{m-1}x_m & \binom{m}{m}x_{m+1}\\ -1 & \binom{m-2}{0}x_1 & \cdots & \binom{m-2}{m-3}x_{m-2} & \binom{m-2}{m-2}x_{m-1}\\ 0 & -1 & \cdots & \binom{m-3}{m-4}x_{m-3} & \binom{m-3}{m-3}x_{m-2}\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & \binom{1}{0}x_1 & \binom{1}{1}x_2\\ 0 & 0 & \cdots & -1 & \binom{0}{0}x_1\end{matrix}\right|.$$

By induction, the first term represents the partitions of $\{1,2,\ldots,m+1\}$ containing $\{m+1\}$ as a singleton block. When we expand the second term across the top row, we get

$$x_1\,\mathbf{Y}_m(x_1,\ldots,x_m)+\sum_{k=1}^{m}\binom{m}{k}x_{k+1}\,c_{m,k}, \tag{2.3}$$

with the same cofactors $c_{m,k}$ as before. For each k between 1 and m, the term $\binom{m}{k}x_{k+1}\,c_{m,k}$ represents all the partitions of $\{1,2,\ldots,m+1\}$ where $m+1$ is in a block with k other elements, since there are $\binom{m}{k}$ ways to choose these elements, and, by induction, $c_{m,k}$ represents all the ways to partition the remaining $m-k$ elements. Therefore (2.3) represents all the partitions of $\{1,2,\ldots,m+1\}$, and hence it must be equal to $\mathbf{Y}_{m+1}(x_1,\ldots,x_{m+1})$, which proves the lemma. Now we are ready for

Faà di Bruno's determinant formula. *If g and f are functions with a sufficient number of derivatives, and $m\geq 1$, then*

$$\frac{d^m}{dt^m}g\,(f(t))$$
$$=\left|\begin{matrix}\binom{m-1}{0}f'g & \binom{m-1}{1}f''g & \binom{m-1}{2}f'''g & \cdots & \binom{m-1}{m-2}f^{(m-1)}g & \binom{m-1}{m-1}f^{(m)}g\\ -1 & \binom{m-2}{0}f'g & \binom{m-2}{1}f''g & \cdots & \binom{m-2}{m-3}f^{(m-2)}g & \binom{m-2}{m-2}f^{(m-1)}g\\ 0 & -1 & \binom{m-3}{0}f'g & \cdots & \binom{m-3}{m-4}f^{(m-3)}g & \binom{m-3}{m-3}f^{(m-2)}g\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & \binom{1}{0}f'g & \binom{1}{1}f''g\\ 0 & 0 & 0 & \cdots & -1 & \binom{0}{0}f'g\end{matrix}\right|,$$

where $f^{(i)}$ denotes $f^{(i)}(t)$ and g^k is to be interpreted as $g^{(k)}(f(t))$.

We need only replace every x_i in the lemma by $f^{(i)}g$ to reduce this determinant to Riordan's formula (2.2), in view of the remark that the $\mathbf{B}_{m,k}$'s may be recovered from the $\mathbf{Y}_m$'s. An equivalent formula was proposed by V. F. Ivanoff in 1958 in this MONTHLY as Advanced Problem #4782. Several years went by before the MONTHLY published a proof (different from ours, and apparently the only one they received) by Frank Schmittroth [**61**]. The editors missed a chance to publicize [**43**], a nice paper on statistical applications of Faà di Bruno's formula that appeared in this MONTHLY three years before the problem was posed, and they also failed to mention Faà di Bruno. (The current editors are, of course, much more enlightened.) Ivanoff's formula is also stated in [**46**], and as a problem in [**12**] and [**13**]. The determinant in [**16**] and [**17**] differs in two ways from the one given here: it is stated for $n+1$, rather than for m; and its first column is our last column, with a compensating factor of $(-1)^n$ in the result. Its columns are arranged as above in all the other occurrences I know of. I have not seen another evaluation—Muir's only comment [**51**] is that "an opportunity was here lost by Bruno of noting that a recurrent with the elements in its zero-bordered diagonal all negative has all its terms positive"—but from the first few pages of [**19**] or [**20**] it seems that Faà di Bruno probably deduced it from two formulas in the theory of symmetric functions.

3. TAYLOR'S THEOREM AND RIVAL FORMULÆ. We may rephrase our initial question: assuming that $f(t)$ and $g(u)$ are sufficiently nice functions, how should we expand $g(f(t+h))$ in powers of h? For Taylor's theorem tells us that

$$g\left(f(t+h)\right)=\sum_{m=0}^{\infty}\frac{d^m}{dt^m}g\left(f(t)\right)\frac{h^m}{m!},\tag{3.1}$$

and, as Ubbo H. Meyer pointed out in 1847 [**48**], it can also be made to say that

$$g\left(f(t+h)\right)=\sum_{k=0}^{\infty}\frac{g^{(k)}(f(t))}{k!}\left(f(t+h)-f(t)\right)^k\tag{3.2}$$

by taking $u=f(t)$ and $v=f(t+h)-f(t)$ in

$$g(u+v)=\sum_{k=0}^{\infty}\frac{g^{(k)}(u)}{k!}v^k.$$

Comparing (3.2) and (3.1) we have

Meyer's observation. $\frac{1}{m!}\frac{d^m}{dt^m}g(f(t))$ *equals the coefficient of* h^m *in (3.2).*

While this idea was not completely new with Meyer, it was never before stated in such a versatile form: several variants of Faà di Bruno's formula follow from different methods of extracting the coefficient of h^m. In [**1**], a few years after [**48**], the pseudonymous T. A. (of whom more later) noted that, if

$$\Delta_h f(t):=\frac{f(t+h)-f(t)}{h},$$

then the coefficient of h^m in $(f(t+h)-f(t))^k$ is the same as that of h^{m-k} in $(\Delta_h f(t))^k$. After taking $m-k$ derivatives with respect to h and setting $h=0$, a little rearrangement gives

T. A.'s formula. *If* g *and* f *are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m}g\left(f(t)\right)=\sum_{k=0}^{m}\binom{m}{k}g^{(k)}(f(t))\left\{\frac{d^{m-k}}{dh^{m-k}}(\Delta_h f(t))^k\right\}\bigg|_{h=0}.$$

Meyer's approach was more straightforward: just take m derivatives of (3.2) with respect to h and set $h = 0$. After cancelling $1/m!$ on each side, and observing that the terms with $k > m$ must vanish, this gives

Meyer's formula. *If g and f are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m} g\,(f(t)) = \sum_{k=0}^{m} \frac{g^{(k)}\,(f(t))}{k!} \left\{ \frac{d^m}{dh^m}\,(f(t+h) - f(t))^k \right\}\bigg|_{h=0}.$$

This result appears in [**37**] under the name "Schlömilch's formula", with reference to [**60**], but Schlömilch attributes it to [**48**] there and in [**59**]. Meyer proceeded to expand $(f(t+h) - f(t))^k$ by the binomial theorem:

$$(f(t+h) - f(t))^k = \sum_{j=0}^{k} \binom{k}{j} (-f(t))^{k-j}\,(f(t+h))^j\,.$$

Now we have only to take m derivatives of $(f(t+h))^j$ with respect to h and then set $h = 0$. Equivalently, by the chain rule, we may take m derivatives of $(f(t+h))^j$ with respect to $t + h$ and then set $h = 0$; but this amounts to taking m derivatives of $(f(t))^j$ with respect to t, and thus Meyer derived

Hoppe's formula. *If g and f are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m} g\,(f(t)) = \sum_{k=0}^{m} \frac{g^{(k)}(f(t))}{k!} A_{m,k}\,(f(t))\,,$$

where

$$A_{m,k}\,(f(t)) = \sum_{j=0}^{k} \binom{k}{j} (-f(t))^{k-j}\,\frac{d^m}{dt^m}\,(f(t))^j\,. \tag{3.3}$$

This also has an easy proof, which I have not seen in the literature, using the recurrence

$$\frac{d}{dt} A_{m,k}(f(t)) = A_{m+1,k}(f(t)) - f'(t) A_{m,k-1}\,(f(t))\,,$$

where $A_{m,0}(f(t)) = 0$ unless $m = 0$, when it equals 1. The theorem was published at least three times by Reinhold Hoppe. It is a main result of his monograph [**32**] of 1845, a summary of which appeared in Crelle's *Journal* the following year [**33**]. I have not seen [**32**], but neither [**33**] nor the much later [**34**] has a proof. If [**32**] has one, we may infer from [**59**] and [**60**] that it is not Meyer's. Meyer seems to have been unaware of Hoppe's work; his only reference is to [**58**], one of two earlier papers of Schlömilch that contains the special cases $f(t) = e^t$ and $f(t) = t^\lambda$ of the interior function. (The similar [**57**] appeared in Crelle's *Journal* in the same year as [**33**].) While Schlömilch does not mention Hoppe's formula in these papers, he did learn of Hoppe's work later, and he gave a different derivation. It is evident that

$$\frac{d^m}{dt^m} g\,(f(t)) = \sum_{i=0}^{m} \frac{g^{(i)}(f(t))}{i!}\,A_{m,i}\,(f(t)) \tag{3.4}$$

for some quantities $A_{m,i}(f(t))$ that depend only on $f(t)$, not on $g(u)$. To find them, take $g(u) = 1, u, u^2, \ldots, u^m$ in turn. Since $d^i\, u^j/du^i = i!\,\binom{j}{i}u^{j-i}$, substituting these choices of $g(u)$ into (3.4) gives

$$\frac{d^m}{dt^m}\,(f(t))^j = \sum_{i=0}^{j}\binom{j}{i}\,(f(t))^{j-i}\,A_{m,i}\,(f(t)) \quad \text{for each } j,\ 0 \le j \le m. \tag{3.5}$$

The rest of the argument can be described succinctly: apply the inverse relation

$$b_j = \sum_{i=0}^{j}\binom{j}{i}(-1)^i\,a_i \quad\Longleftrightarrow\quad a_k = \sum_{j=0}^{k}\binom{k}{j}(-1)^j\,b_j$$

to (3.5). Schlömilch did not cast it in these terms, but presumably one would not find this proof without some inkling of the general principle. Multiply (3.5) by $\binom{k}{j}(-f(t))^{-j}$ for each j, $0 \le j \le k$, and sum on j:

$$\begin{aligned}
&\sum_{j=0}^{k}\binom{k}{j}\,(-f(t))^{-j}\,\frac{d^m}{dt^m}\,(f(t))^j \\
&\quad = \sum_{j=0}^{k}\binom{k}{j}\,(-f(t))^{-j}\sum_{i=0}^{j}\binom{j}{i}\,(f(t))^{j-i}\,A_{m,i}\,(f(t)) \\
&\quad = \sum_{i=0}^{k}\binom{k}{i}\,(-f(t))^{-i}\,A_{m,i}\,(f(t))\sum_{j=i}^{k}\binom{k-i}{j-i}(-1)^{j-i}.
\end{aligned}$$

The inner sum is $(1-1)^{k-i}$, so it equals 1 if $i = k$ and is 0 otherwise. Then we have

$$\sum_{j=0}^{k}\binom{k}{j}\,(-f(t))^{-j}\,\frac{d^m}{dt^m}\,(f(t))^j = (-f(t))^{-k}\,A_{m,k}\,(f(t)),$$

which is (3.3). Schlömilch gave this argument in [**59**], and again later in [**60**]. An interesting variation was given by Mossa [**49**]. First note the

Lemma. *If $f_1, f_2, \ldots, f_k$ are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m}\,\{f_1(t)\,f_2(t)\cdots f_k(t)\} = \sum_{j_1+\cdots+j_k=m}\binom{m}{j_1,\ldots,j_k}f_1^{(j_1)}(t)\cdots f_k^{(j_k)}(t).$$

Taking each $f_i(t)$ equal to $f(t)$ here gives

$$\frac{d^m}{dt^m}\,(f(t))^k = \sum_{j_1+\cdots+j_k=m}\binom{m}{j_1,\ldots,j_k}f^{(j_1)}(t)\cdots f^{(j_k)}(t), \tag{3.6}$$

and comparing (3.6) with (3.5) we get

$$\sum_{i=0}^{k}\binom{k}{i}\,(f(t))^{k-i}\,A_{m,i}\,(f(t)) = \sum_{j_1+\cdots+j_k=m}\binom{m}{j_1,\ldots,j_k}f^{(j_1)}(t)\cdots f^{(j_k)}(t). \tag{3.7}$$

Mossa then asked for the coefficient of $(f(t))^{k-i}$ on the right side of (3.7). To get $(f(t))^{k-i}$ there we need exactly $k-i$ of the j_h's to be zero. There are $\binom{k}{i}$ different ways to choose the positive j_h's, so the desired coefficient is

$$\binom{k}{i} \sum_{\substack{j_1+\cdots+j_i=m \\ j_h \geq 1}} \binom{m}{j_1, \ldots, j_i} f^{(j_1)}(t) \cdots f^{(j_i)}(t),$$

and therefore (3.7) implies that

$$A_{m,i}(f(t)) = \sum_{\substack{j_1+\cdots+j_i=m \\ j_h \geq 1}} \binom{m}{j_1, \ldots, j_i} f^{(j_1)}(t) \cdots f^{(j_i)}(t)$$

if $m \geq 1$; this, together with (3.4), is Riordan's formula (2.2).

Schlömilch's argument also appears, without any reference, in Bertrand's *Traité de Calcul différentiel* [**6**] of 1864, which seems to be the first appearance of Hoppe's formula in the French literature. (Although Meyer's paper [**48**] is in French, it appeared in Grunert's *Archiv der Mathematik und Physik.* Coincidentally, 25 years later Hoppe became editor of the *Archiv* when Grunert died in 1872, and held the job until his own passing in 1900. Thus he had no lack of opportunities to publish any proof that he might have possessed.) Bertrand pointed out that we may rewrite (3.3) as

$$A_{m,k}(f(t)) = (f(t))^k \frac{d^m}{dt^m} \left(\frac{f(t)}{\alpha} - 1 \right)^k, \tag{3.8}$$

where α is to be considered a constant until the binomial has been expanded and the differentiations performed, and only then set equal to $f(t)$. Fais also made this observation [**21**], though he may not have done so independently. Hoppe's formula is discussed on pp. 138–141 of [**6**], and Faà di Bruno's formula (without a reference) on pp. 308–309. Fais gives a reference to [**6**] for the latter, but not the former, so he might have missed (3.8), but it could also be that he forgot he had seen it there. Hermite included Hoppe's formula (but not Faà di Bruno's) in his *Cours d'Analyse* [**30**], attributing it to [**6**] while giving Meyer's proof.

Hoppe's formula could be considered an unsatisfactory solution to the higher chain rule problem, in that it merely reduces to the case where the exterior function g is a power function. A more interesting remark is that no powers of f itself (nor any minus signs) can actually appear in the final answer. This means that all the terms in (3.3) with $j < k$ must ultimately cancel—the only contributions that survive come from $j = k$. Some constituents of the $j = k$ term still have a factor $f(t)$, so they too must vanish; thus Hoppe's formula simplifies to

Scott's formula. *If g and f are functions with a sufficient number of derivatives, then*

$$\frac{d^m}{dt^m} g(f(t)) = \sum_{k=0}^{m} \frac{g^{(k)}(f(t))}{k!} \left\{ \left[\frac{d^m}{dt^m} (f(t))^k \right] \Bigg|_{f(t)=0} \right\}.$$

This theorem (more or less) is in a little-known paper from 1861 by George Scott [**63**], which is, I believe, the first work to contain both Faà di Bruno's and Hoppe's formulas, and the first appearance of either in English. Both are unreferenced, but it seems reasonable to suppose that Scott was at least aware of [**17**], which was in the *Quarterly Journal of Pure and Applied Mathematics* just four years before [**63**] appeared there. I know two other sources in English for Hoppe's formula: Schwatt's book [**62**], with no reference, where one of the proofs is a less elegant version of Schlömilch's; and [**23**] (whose title deserves some sort of award), with a reference to [**6**].

If we set $f(t) = 0$ in (3.6) (so that the only terms that survive are those with each $j_i \geq 1$) and substitute in Scott's formula, Riordan's formula falls out. This remark is essentially due to Marchand, in his survey paper [**45**], the only other reference I know for Scott's formula (which one could also prove by using Marchand's idea in the other direction). Fais proved Faà di Bruno's formula in a similar way using (3.4), (3.8), and (3.6) [**21**]. One may also derive Faà di Bruno's formula directly from (3.2):

$$\begin{aligned}(f(t+h) - f(t))^k &= \left(\sum_{j_1=1}^{\infty} f^{(j_1)}(t)\frac{h^{j_1}}{j_1!}\right) \cdots \left(\sum_{j_k=1}^{\infty} f^{(j_k)}(t)\frac{h^{j_k}}{j_k!}\right) \\ &= \sum_{m=k}^{\infty} \frac{h^m}{m!} \sum_{\substack{j_1+\cdots+j_k=m \\ j_i \geq 1}} \binom{m}{j_1, \ldots, j_k} f^{(j_1)}(t) \cdots f^{(j_k)}(t) \\ &= k! \sum_{m=k}^{\infty} \frac{h^m}{m!} \mathbf{B}_{m,k}\left(f'(t), f''(t), \ldots, f^{(m-k+1)}(t)\right),\end{aligned}$$

and so (3.2) becomes

$$\begin{aligned}g(f(t+h)) &= \sum_{k=0}^{\infty} g^{(k)}(f(t)) \sum_{m=k}^{\infty} \frac{h^m}{m!} \mathbf{B}_{m,k}\left(f'(t), f''(t), \ldots, f^{(m-k+1)}(t)\right) \\ &= \sum_{m=0}^{\infty} \frac{h^m}{m!} \sum_{k=0}^{m} g^{(k)}(f(t))\, \mathbf{B}_{m,k}\left(f'(t), f''(t), \ldots, f^{(m-k+1)}(t)\right).\end{aligned}$$

Here is an expansion of $g(f(t+h))$ in powers of h, and Riordan's formula follows after comparison with (3.1). Bertrand obtained the classical Faà di Bruno formula (with a few misprints) by this method in [**6**]. His argument appears in the German literature much later, in a paper by Franz Meyer [**47**], who seems to have learned it from Dedekind. The proof in [**63**] is also along these lines, using the symbolic form $\phi(y+h) = e^{h\frac{d}{dy}}\phi(y)$ of Taylor's theorem. The main idea is older yet, as we see in the next section.

4. THE SECRET HISTORY OF FAÀ DI BRUNO'S FORMULA. Schlömilch's proof of Hoppe's formula exploits the fact that, since one can see *a priori* that the factor multiplying $g^{(k)}(f(t))$ depends only on f, not on g, one may specialize g to find it. The best-known proof of Faà di Bruno's formula also uses this idea, with an exponential function for g. Ironically, it dates back to a paper that is almost entirely unknown, namely [**1**], which we met in the preceding section. Five years before Faà di Bruno's papers, T. A. expanded $e^{p\varphi(x+h)}$ in powers of h. On one hand, by Taylor's theorem,

$$e^{p\varphi(x+h)} = \sum_{n=0}^{\infty} \left\{\frac{d^n}{dx^n} e^{p\varphi(x)}\right\} \frac{h^n}{n!}. \tag{4.1}$$

On the other hand, if we expand $\varphi(x+h)$ then we get

$$e^{p\varphi(x+h)} = e^{p\varphi(x)} e^{p\varphi'(x)h} e^{p\varphi''(x)\frac{h^2}{2!}} \cdots$$

Developing all factors except the first in infinite series and rearranging, we have

$$e^{p\varphi(x+h)} = \sum_{n=0}^{\infty} h^n \sum_{k=0}^{n} p^k e^{p\varphi(x)} \sum \frac{1}{b_1! \cdots b_n!} \left(\frac{\varphi'(x)}{1!}\right)^{b_1} \cdots \left(\frac{\varphi^{(n)}(x)}{n!}\right)^{b_n}, \quad (4.2)$$

where the innermost sum is over all different collections of nonnegative integers $b_1, \ldots, b_n$ satisfying $b_1 + b_2 + \cdots + b_n = k$ and $b_1 + 2b_2 + \cdots + nb_n = n$. T. A. obtained what we now call Faà di Bruno's formula by comparing (4.1) with (4.2).

T. A. was described in [**1**] and its sequel [**2**] as "Ancien élève [alumnus] de l'École Polytechnique". It appears that he was an artillery captain named J. F. C. Tiburce Abadie. He made several contributions to the *Nouvelles Annales de Mathématiques* around 1850, of which (in my opinion) [**1**] and [**2**] are the most notable. He seems to have stopped doing mathematics soon afterwards, as his last work for the *Nouvelles Annales* came out in 1855. Apart from [**2**], I know of only three references to [**1**]. Fais and Mossa mentioned it in the same volume of Battaglini's *Giornale* in 1875 [**21**], [**49**], and Marchand discussed [**1**] and [**2**] in [**45**] a decade later. [**2**] was one of several papers appended by Combescure to [**8**], his French translation of Brioschi's book on determinants, and Muir also comments on it in [**51**]. It begins with T. A.'s formula of the preceding section, and never mentions Faà di Bruno's formula.

There are several umbral calculus proofs in the same spirit as T. A.'s proof. The first, by Riordan [**53**], predates the rigorization of umbral techniques by Rota and several collaborators. Subsequent papers by Roman [**54**] and Chen [**11**] eventually put Riordan's argument on a firm foundation.

The first doctoral thesis in mathematics at the Friedrich-Wilhelms-Universität in Berlin (see [**7**]) was [**55**], completed by Heinrich Ferdinand Scherk in 1823. In section 5 of his dissertation, Scherk wrote that, if z is a function of y and y a function of x, then (in his notation) the coefficients $\overset{k}{A}$ in the expansion

$$\frac{d^n z}{dx^n} = \overset{n}{A} \frac{d^n z}{dy^n} + \overset{n-1}{A} \frac{d^{n-1} z}{dy^{n-1}} + \overset{n-2}{A} \frac{d^{n-2} z}{dy^{n-2}} + \cdots + \overset{k}{A} \frac{d^k z}{dy^k} + \cdots + \overset{1}{A} \frac{dz}{dy} \quad (4.3)$$

are

$$\overset{k}{A} = \sum \frac{\left(\frac{dy}{dx}\right)^{\overset{1}{\alpha}} \left(\frac{d^2 y}{dx^2}\right)^{\overset{2}{\alpha}} \left(\frac{d^3 y}{dx^3}\right)^{\overset{3}{\alpha}} \cdots \left(\frac{d^{n-k+1} y}{dx^{n-k+1}}\right)^{\overset{n-k+1}{\alpha}}}{(\prod 1)^{\overset{1}{\alpha}} (\prod 2)^{\overset{2}{\alpha}} (\prod 3)^{\overset{3}{\alpha}} \cdots [\prod (n-k+1)]^{\overset{n-k+1}{\alpha}}} \frac{\prod n}{\prod^{1}_{\alpha} \cdot \prod^{2}_{\alpha} \cdot \prod^{3}_{\alpha} \cdots \prod^{n-k+1}_{\alpha}}, \quad (4.4)$$

where the sum is over all collections of nonnegative integers $\overset{1}{\alpha}, \overset{2}{\alpha}, \overset{3}{\alpha}, \ldots, \overset{n-k+1}{\alpha}$ satisfying the pair of equations

$$\begin{aligned} \overset{1}{\alpha} + 2\overset{2}{\alpha} + 3\overset{3}{\alpha} + \cdots + (n-k+1)\overset{n-k+1}{\alpha} &= n, \\ \overset{1}{\alpha} + \overset{2}{\alpha} + \overset{3}{\alpha} + \cdots + \overset{n-k+1}{\alpha} &= k. \end{aligned}$$

Scherk's formulation incorporates the fact that there can be at most $n - k + 1$ positive α's, as in the definition of the Bell polynomials, a reduction not made by most later writers. He pointed out that Crelle had recently considered the same problem in [**15**]. Crelle gave a number of special cases of Faà di Bruno's formula, up to $n = 6$; there is also an unsuccessful attempt to write down the general case. In some sense he knew what the formula should be, but could not see how to describe it in general. This may explain his interest in Hoppe's work many years later.

Scherk's evident motivation is rather surprising. His thesis is primarily concerned with Stirling numbers. After writing down an example with $n = 5$ and $k = 3$, he observed that if we take $y = e^x$ and $z = e^{ax} = y^a$ in (4.3) and (4.4), then we can derive

$$a^n = \sum_{k=0}^{n} \left\{ {n \atop k} \right\} (-r)^{n-k} (a)_{k,r}, \tag{4.5}$$

which he attributed to Kramp's book [**40**]; this expresses the power functions a^n in terms of the rising factorials $(a)_{k,r} := a(a+r)(a+2r)\dots(a+(k-1)r)$, where $(a)_{0,r} := 1$. To be more precise, with these choices of y and z, the sum for $\overset{k}{A}$ may be evaluated by (2.1), and this results in the case $r = -1$ of (4.5). The general case then follows by replacing a by $-a/r$. One might instead prove (4.5) by induction, using the recurrence

$$\left\{ {n+1 \atop k} \right\} = \left\{ {n \atop k-1} \right\} + k \left\{ {n \atop k} \right\}.$$

Several papers on Faà di Bruno's formula refer to p. 325 of the first volume of Lacroix's *Traité du Calcul différentiel et du Calcul intégral*, 2nd edition [**41**], which dates back to 1810. Lacroix wrote there that, if y is a function of x, then

$$\begin{aligned} &\phi(y) + \frac{d\phi(y)}{dx}\frac{dx}{1} + \frac{d^2\phi(y)}{dx^2}\frac{dx^2}{1.2} + \frac{d^3\phi(y)}{dx^3}\frac{dx^3}{1.2.3} + \text{etc.} \\ &= \phi\left(y + \frac{dy}{dx}\frac{dx}{1} + \frac{d^2y}{dx^2}\frac{dx^2}{1.2} + \frac{d^3y}{dx^3}\frac{dx^3}{1.2.3} + \text{etc.}\right) \end{aligned} \tag{4.6}$$

Lacroix concluded that $d^n\phi(y)/1.2\dots n\,dx^n$ must equal the coefficient of dx^n on the right side (a precursor of Meyer's observation), but volume 1 leaves it at that. Volume 3 came out nine years later, and by then Lacroix had seen how to complete the calculation. Assuming again that y is a function of x, he wrote

$$\begin{aligned} \frac{d^n\phi(y)}{dx^n} = \frac{d^n\phi(y)}{dy^n}T_0^n + n\frac{d^{n-1}\phi(y)}{dy^{n-1}}T_1^{n-1} + n(n-1)\frac{d^{n-2}\phi(y)}{dy^{n-2}}T_2^{n-2}\dots\dots\dots\dots\dots \\ \dots\dots\dots + n(n-1)(n-2)\dots.2\frac{d\phi(y)}{dy}T_{n-1}^1, \end{aligned}$$

where T_s^r is the coefficient of dx^s in the expansion of

$$\left(\frac{dy}{dx} + \frac{d^2y}{dx^2}\frac{dx}{1.2} + \frac{d^3y}{dx^3}\frac{dx^2}{1.2.3} + \text{etc.}\right)^r,$$

which, he further wrote (with reference to section 24 of the *Introduction* in volume 1), is

$$\frac{1.2.3.\dots r}{1.2\dots a.1.2\dots b.1.2\dots c \times \text{etc.}} \times \frac{\left(\frac{dy}{dx}\right)^a \left(\frac{d^2y}{dx^2}\right)^b \left(\frac{d^3y}{dx^3}\right)^c \text{ etc.}}{(1)^a(1.2)^b(1.2.3)^c \text{ etc.}},$$

where $a, b, c, \ldots$ satisfy

$$\begin{aligned} a + b + c + \text{etc.} &= r, \\ b + 2c + \text{etc.} &= s. \end{aligned}$$

This is on p. 629 of volume 3 of [**41**], in a long concluding section entitled *Corrections et Additions*. It is not in the first edition of the *Traité*, whose three volumes were published in 1797, 1798, and 1800, and (4.6) is not there either.

[**41**] is a snapshot (or perhaps a mural) of 18th century analysis, especially valuable now for the extensive bibliography in most sections. Although he was the outstanding textbook author of his time, it has been written that Lacroix made no original contribution to mathematics (on this point see [**29**, p. 113]), so one naturally wonders whether he had a source for this material. He does not give a reference for Faà di Bruno's formula. The section in which (4.6) appears [**41**, vol. 1, pp. 315–326] lists three sources, one of which is *Du Calcul des Dérivations*, published in 1800 by L. F. A. Arbogast [**4**]. Much of the section is devoted to Arbogast's work. While (4.6) comes near the end, and is not specifically attributed to Arbogast, on p. 3 of [**4**] one finds

$$\phi\left(a + \frac{D.a}{1}x + \frac{D^2.a}{1.2}x^2 + \text{etc.}\right) = \phi a + \frac{D.\phi a}{1}x + \frac{D^2.\phi a}{1.2}x^2 + \text{etc.}$$

Arbogast noted that this equation could be used to work out the higher derivatives $D^n.\phi a$, giving examples when $n = 2$ and $n = 3$.

The first six cases of Faà di Bruno's formula are on pp. 310–311 of [**4**]. At the bottom of p. 312 Arbogast gives a prose rule for writing down the general case, and on p. 313 he illustrates his rule by writing down the case $n = 6$ again, twice, followed by $n = 7$. Although he never gives a general formula, nor a proof of his rule, on pp. 43–44 of [**4**] there is a formula for the coefficient of x^m in $(\zeta + \gamma x + \delta x^2 + \epsilon x^3 + \text{etc.})^n$. Thus Arbogast had most of the ingredients of Lacroix's argument at hand, but he seems never to have written down Faà di Bruno's formula as such.

5. CONCLUSION. One occasionally sees the variant spelling "Faà de Bruno", which Faà di Bruno himself always used when writing in French. In his Italian papers he seems to be invariably "Cav. F. Faà di Bruno". Francesco is his first name, and "Cav." signifies "Cavaliere"; he was in the army for a number of years before deciding to study mathematics. In French he is most often "M. Faà de Bruno", but in [**19**] he is "Chevalier François Faà de Bruno", and in [**20**] "Chev. F. Faà de Bruno".

Another bit of trivia: although Scherk published Faà di Bruno's formula in Berlin in 1823, his thesis is in Latin, and the formula was apparently never printed in German before 1871. Hoppe restated his formula in 1870 [**34**], in volume 4 of the *Mathematische Annalen*, in response to the publication by Götting in volume 3 of a complementary result: the special case of Faà di Bruno's formula where $g(u) = u^k$ [**27**]. Volume 4 also has Most's paper [**50**], which gives the general case. Scherk was still alive then, but, in his seventies and not mathematically active, he did not follow Hoppe's example.

Scherk and Crelle were both familiar with Lacroix's book. While Scherk did not mention it in his thesis [**55**], in 1825 he published a short book [**56**] which has several references to [**41**]. Some of the material in [**55**] is repeated in [**56**], but Faà di Bruno's formula is not; thus Scherk missed an opportunity to publicize it, just as T. A. did 27 years later. There are several plausible reasons for this omission: Scherk may have known that the formula was in [**41**] (I think he probably did not know this in 1823, but he may have known it by 1825); he may have been bothered by a lack of proof; or he may simply have thought it wasn't that interesting. The first reason could also apply to T. A.

Although [**15**] contains several references to [**41**], Crelle must not have known that Faà di Bruno's formula was there, or else he would have been able to write down the general case himself. How could he, and others, have missed it? Some contributing factors may be:

i. It was only in the second edition of the *Traité*, not the first.
ii. It was not in the second edition where it logically belonged, in the middle of the first volume, but rather near the end of the last volume, in a supplement.
iii. It was not the *Traité* but an abridgement, the *Traité Elementaire*, which was more commonly used as a textbook.
iv. The translations of the *Traité* into German and into English were also abridged.
v. At the time that the last volume of [**41**] came out, Cauchy was reforming the calculus sequence at the École Polytechnique. The impact of this was profound; it is still felt today in calculus and analysis, and over time it diminished the influence of the *Traité* considerably.

We could add to this list the fact that Bertrand's *Traité* was in many ways a new and improved version of Lacroix's, so whatever audience [**41**] still had by 1864 was further reduced by [**6**]. Bertrand may be the most likely of our authors to have known that Faà di Bruno's formula was in [**41**]. I also think he is as likely to have read [**1**] and [**2**] as [**16**] or [**17**].

Faà di Bruno's other mathematical accomplishments, of which an excellent account may be found in [**9**], have caused him to receive more credit for the higher chain rule than he deserves. His book [**20**] on binary forms was well known throughout Europe—written in French, published in Italy, translated into German, and on a subject of great interest to Cayley and Sylvester, leading British mathematicians of the time. The name "Faà di Bruno's formula" came into use only after the appearance of [**20**]. As far as I know, the only references before that to Faà di Bruno's work in this area are to [**16**] by Italians (Faà di Bruno himself and [**22**], [**21**], [**49**]).

I cannot explain how Faà di Bruno, who was living in Paris when he wrote [**16**], [**17**], and [**18**], could have overlooked [**1**], which appeared in a prominent French journal only five years earlier. I find this one of the most puzzling aspects of the whole subject, especially since the sequel [**2**] was in the same journal only three years before Faà di Bruno wrote [**16**] and [**17**]. One wonders about his decision to publish the French version [**17**] in a British journal, which made it less likely that anyone familiar with [**1**] or [**41**] would see it. But there is a very good reason why someone whose primary interest was invariant theory should have sent a paper to the *Quarterly Journal of Pure and Applied Mathematics* at that time: Sylvester had just become its editor.

One also wonders how two Italians knew of [**1**] when it seems to have been virtually unknown in France. Perhaps Fais and Mossa (who was evidently Fais' student) got this reference from Faà di Bruno—who else in Italy in 1875 was more likely to know about T. A.'s work? On the other hand, if we are to judge only from their publications, Fais seems more familiar with the literature in this area than Faà di Bruno; he also cites [**27**] and [**50**]. One might instead wonder whether Faà di Bruno had read [**21**] or [**49**] by the time he completed [**20**].

It is easy to see why Faà di Bruno's formula should have won out over Hoppe's formula and the other candidates, and one might better ask why Hoppe's formula was seen as a serious rival in the 1860s and 1870s. I believe that the Bell polynomial (or equivalently the set partition) version is really the fundamental form of the result.

While the demise of Hoppe's formula is unlamented, I also think that Scott's formula deserves to be better known.

Finally, one is struck by the obscurity of many of the names in our story. (An easy if not altogether satisfying explanation: this is made possible by the lack of depth of the subject.) It would be nice to have more biographical information about some of them, T. A. especially. I know of two good sources for Hoppe: the obituary notice [**42**], and Biermann's history [**7**] of the mathematics department at Berlin University. The latter also has some information about Scherk. Background on Lacroix may be found in [**29**], an unsurpassed reference for early 19th century French mathematics. Faà di Bruno has been the subject of several Italian biographies focusing on the religious aspects of his life. He was ordained a Roman Catholic priest on October 22, 1876, was a pioneer in charitable works, and was declared a Saint on September 25, 1988. In his last two papers for Sylvester's *American Journal of Mathematics* he was "the Rev. Faà de Bruno" and "l'Abbé Faà de Bruno". One might only wish that this admirable man had either read more or been a little more saintly in his citations.

6. ACKNOWLEDGEMENTS AND SOURCES. I seldom thought about Faà di Bruno's formula after [**36**] was published until I saw a preliminary version of [**68**], which rekindled my interest. I thank Winston Yang for writing it, and George Andrews for sending it to me.

I should perhaps say something further about references. I might never have begun this project had I not come across [**45**] and the reference there to T. A.'s work; thus I owe much to [**13**] and the outstanding bibliography therein. Marchand seems to have read nearly everything ever published on this subject in France, but not much else. Unfortunately he failed to spot Faà di Bruno's formula in [**41**]. He also missed [**17**] and [**48**], which were both written in French but published elsewhere.

Although I was misled by it once or twice, a short history of the higher chain rule that was very helpful is in Lukacs' paper [**43**]. Much of Lukacs' history was evidently drawn from the article by Voss in the *Encyklopädie der Mathematischen Wissenschaften* [**67**]. Voss is an excellent source on the German contributions to the subject from 1845 on; otherwise he mentions only [**16**] and [**41**], though he too missed Faà di Bruno's formula in Lacroix's third volume. To paraphrase an old joke, Voss is Marchand with some changes of sign.

REFERENCES

1. T. A. [J. F. C. Tiburce Abadie], Sur la différentiation des fonctions de fonctions, *Nouvelles Annales de Mathématiques* **9** (1850) 119–125.
2. A. [J. F. C. Tiburce Abadie], Sur la différentiation des fonctions de fonctions. Séries de Burmann, de Lagrange, de Wronski, *Nouvelles Annales de Mathématiques* **11** (1852) 376–383.
3. George E. Andrews, The theory of partitions, *Encyclopedia of Mathematics and its Applications*, vol. 2, Addison-Wesley Publishing Company, Reading, Massachusetts, 1976; Cambridge University Press, Cambridge, 1984, Cambridge Mathematical Library, 1998.
4. L. F. A. Arbogast, *Du Calcul des Dérivations*, Levrault, Strasbourg, 1800.
5. Eric Temple Bell, Exponential polynomials, *Ann. of Math. (2)* **35** (1934) 258–277.
6. Joseph Bertrand, *Traité de Calcul différentiel et de Calcul intégral*, vol. 1, Gauthier-Villars, Paris, 1864.
7. Kurt-R. Biermann, *Die Mathematik und ihre Dozenten an der Berliner Universität 1810–1933*, Akademie-Verlag, Berlin, 1988.
8. Francesco Brioschi, *Theorie des déterminants et leurs principales applications*; French translation by Edouard Combescure, Mallet-Bachelier, Paris, 1856.
9. Giuseppina Casadio and Guido Zappa, I contributi matematici di Francesco Faà di Bruno nel periodo 1873–1881, con particolare riguardo alla teoria degli invarianti, *Rend. Circ. Mat. Palermo (2) Suppl.* **36** (1994) 47–70.
10. Ernesto Cesàro, Dériveés des fonctions de fonctions, *Nouvelles Annales de Mathématiques, 3rd series* **4** (1885) 41–55; Opere Scelte, vol. 1, Edizioni Cremorese, Rome, 1964, pp. 416–429.

11. William Y. C. Chen, Context-free grammars, differential operators and formal power series, Conference on Formal Power Series and Algebraic Combinatorics (Bordeaux, 1991), *Theoret. Comput. Sci.* **117** (1993) 113–129.
12. Louis Comtet, *Analyse Combinatoire*, Presses Universitaires de France, Paris, 1970.
13. Louis Comtet, *Advanced Combinatorics*, Revised and Enlarged Edition, D. Reidel Publishing Company, Dordrecht, Holland, 1974.
14. G. M. Constantine, *Combinatorial Theory and Statistical Design*, John Wiley & Sons, New York, 1987.
15. August Leopold Crelle, *Sammlung mathematischer Aufsätze und Bemerkungen*, vol. 2, Maurer, Berlin, 1822.
16. Cavaliere Francesco Faà di Bruno, Sullo sviluppo delle Funzioni, *Annali di Scienze Matematiche e Fisiche* **6** (1855) 479–480.
17. Cavaliere Francesco Faà di Bruno, Note sur une nouvelle formule de calcul différentiel, *Quarterly J. Pure Appl. Math.* **1** (1857) 359–360.
18. Cavaliere Francesco Faà di Bruno, Sulle Funzioni Isobariche, *Annali di Scienze Matematiche e Fisiche* **7** (1856) 76–89.
19. Cavaliere Francesco Faà di Bruno, *Théorie générale de l'élimination*, Lieber et Faraguet, Paris, 1859.
20. Cavaliere Francesco Faà di Bruno, *Theorie des formes binaires*, librarie Brero, Torino, 1876.
21. Antonio Fais, Nota intorno alle derivate d'ordine superiore delle funzioni di funzione, *Giornale di Matematiche* **13** (1875) 47–48.
22. Emmanuele Fergola, Sopra due formule di calcolo differenziale, *Annali di Matematica Pura ed Applicata* **1** (1858) 370–378.
23. J. C. Fields, The expression of any differential coefficient of a function of a function of any number of variables by aid of the corresponding differential coefficient of any n powers of the function, where n is the order of the differential coefficient, *Amer. J. Math.* **11** (1889) 388–396.
24. Harley Flanders, From Ford to Faà, this MONTHLY **108** (2001) 559–561.
25. Roberto Frucht, A combinatorial approach to the Bell polynomials and their generalizations, in *Recent Progress in Combinatorics*, W. T. Tutte, ed., Proceedings of the Third Waterloo Conference on Combinatorics, May 1968, Academic Press, New York, 1969, pp. 69–74.
26. Roberto Frucht and Gian-Carlo Rota, Polinomios de Bell y particiones de conjuntos finitos, *Scientia* **126** (1965) 5–10.
27. R. Götting, Differentiation des Ausdruckes x^k wenn x eine Funktion irgend einer unabhänging Veränderlichen bedeutet, *Math. Ann.* **3** (1870) 276–285.
28. Édouard Goursat, *Cours d'Analyse Mathématique*, vol. 1, Gauthier-Villars, Paris, 1902.
29. Ivor Grattan-Guinness, *Convolutions in French mathematics 1800–1840*, vol. 1, Birkhäuser Verlag, Basel, 1990.
30. Charles Hermite, *Cours d'Analyse de l'École Polytechnique*, Gauthier-Villars, Paris, 1873.
31. E. Hess, Zur Theorie der Vertauschung der unabhängigen Variablen, *Zeitschrift für Mathematik und Physik* **17** (1872) 1–12.
32. Reinhold Hoppe, *Theorie der independenten Darstellung der höheren Differentialquotienten*, Johann Ambrosius Barth, Leipzig, 1845.
33. Reinhold Hoppe, Ueber independente Darstellung der höheren Differentialquotienten und den Gebrauch des Summenzeichens, *J. Reine Angew. Math.* **33** (1846) 78–89.
34. Reinhold Hoppe, Ueber independente Darstellung der höheren Differentialquotienten, *Math. Ann.* **4** (1871) 85–87.
35. Roger A. Horn and Charles R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
36. Warren P. Johnson, A q-analogue of Faà di Bruno's formula, *J. Combin. Theory Ser. A* **76** (2) (1996) 305–314.
37. Charles Jordan, *Calculus of Finite Differences*, 2nd ed., Chelsea, New York, 1950.
38. Donald E. Knuth, *The Art of Computer Programming, Volume 1. Fundamental Algorithms*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1968.
39. Donald E. Knuth, Two notes on notation, this MONTHLY **99** (1992) 403–422.
40. Christian Kramp, *Analyse des rèfractions astronomiques et terrestres*, P. J. Dannbach, Strasbourg, 1799.
41. Silvestre-François Lacroix, *Traité du Calcul différentiel et du Calcul intégral*, 2e Édition, Revue et Augmentée, vol. 1, Chez Courcier, Paris, 1810; vol. 2, Mme. Courcier, Paris, 1814; vol. 3, Mme. Courcier, Paris, 1819.
42. Emil Lampe, Nachruf für Reinhold Hoppe, *Archiv der Mathematik und Physik, Generalregister zu den Bänden 1–17 der zweiten Reihe, (1884–1900), 1901, VII–XXII.*
43. Eugene Lukacs, Applications of Faà di Bruno's formula in mathematical statistics, this MONTHLY **62** (1955) 340–348.

44. Ian G. Macdonald, *Symmetric Functions and Hall Polynomials*, 2nd ed., Oxford Mathematical Monographs, Oxford University Press, Oxford, 1995.
45. E. Marchand, Sur le changement de variables, *Annales scientifiques de l'école normale supérieure, 3rd series* **3** (1886) 137–188.
46. Z. A. Melzak, *Companion to Concrete Mathematics*, Wiley-Interscience, New York, 1973.
47. Franz Meyer, Ueber algebraische Relationen zwischen den Entwickelungscoefficienten höherer Differentiale, *Math. Ann.* **36** (1890) 453–466.
48. Ubbo H. Meyer, Sur les dériveés d'une fonction de fonction, *Archiv der Mathematik und Physik* **9** (1847) 96–100.
49. Francesco Mossa, Sulla derivazione successiva delle funzioni composte, *Giornale di Matematiche* **13** (1875) 175–185.
50. R. Most, Ueber die höheren Differentialquotienten, *Math. Ann.* **4** (1871) 499–504.
51. Sir Thomas Muir, *The Theory of Determinants in the Historical Order of Development*, vol. 2, Macmillan and Co., Limited, London, 1923.
52. John Riordan, Derivatives of composite functions, *Bull. Amer. Math. Soc.* **52** (1946) 664–667.
53. John Riordan, *An Introduction to Combinatorial Analysis*, John Wiley & Sons, New York, 1958; Princeton University Press, Princeton, NJ, 1980.
54. Steven Roman, The formula of Faà di Bruno, this MONTHLY **87** (1980) 805–809.
55. Heinrich Ferdinand Scherk, *De evolvenda functione* $\frac{yd.\,yd.\,yd...ydX}{dx^n}$ *disquisitiones nonnullae analyticae*, Dissertation, Friedrich-Wilhelms-Universität, Berlin, 1823.
56. Heinrich Ferdinand Scherk, *Mathematische Abhandlungen*, G. Reimer, Berlin, 1825.
57. Oskar Schlömilch, Théorèmes généraux sur les dérivées d'un ordre quelconque de certaines fonctions très générales, *J. Reine Angew. Math.* **32** (1846) 1–7.
58. Oskar Schlömilch, Allgemeine Sätze für eine Theorie der höheren Differential-Quotienten, *Archiv der Mathematik und Physik* **7** (1846) 204–214.
59. Oskar Schlömilch, Zur Theorie der höheren Differentialquotienten, *Zeitschrift für Mathematik und Physik* **3** (1858) 65–80.
60. Oskar Schlömilch, *Compendium der Höheren Analysis*, vol. 2, Friedrich Vieweg und Sohn, Braunschweig, 1866.
61. Frank Schmittroth, Solution to advanced problem #4782, this MONTHLY **68** (1961) 69–70.
62. I. J. Schwatt, *An Introduction to the Operations with Series*, The Press of the University of Pennsylvania, Philadelphia, 1924.
63. George Scott, Formulæ of successive differentiation, *Quarterly J. Pure Appl. Math.* **4** (1861) 77–92.
64. Richard P. Stanley, *Enumerative Combinatorics, Volume 1*, Cambridge Studies in Advanced Mathematics 49, Cambridge University Press, Cambridge, 1997.
65. Richard P. Stanley, *Enumerative Combinatorics, Volume 2*, Cambridge Studies in Advanced Mathematics 62, Cambridge University Press, Cambridge, 1999.
66. Charle Jean de la Vallée Poussin, *Cours d'Analyse Infinitésimale*, 3rd ed., vol. 1, Gauthier-Villars, Paris, 1914.
67. A. Voss, Differential- und Integralrechnung, in *Encyklopädie der Mathematischen Wissenschaften*, II/1/1, B. G. Teubner, Leipzig, 1899, pp. 54–134.
68. Winston C. Yang, Derivatives are essentially integer partitions, *Discrete Math.* **222** (2000) 235–245.
69. Doron Zeilberger, Toward a combinatorial proof of the Jacobian conjecture?, in *Combinatoire énumérative*, Lecture Notes in Mathematics 1234, Springer-Verlag, Berlin, 1986, pp. 370–380.

WARREN P. JOHNSON was an undergraduate at the University of Minnesota. He received his Ph.D. from the University of Wisconsin under the direction of Richard Askey. He has taught at Penn State University, Beloit College, and the University of Wisconsin, and is now at Bates College. He always enjoyed the historical notes in Askey's special functions courses, and this paper satisfies the vague ambition he had of discovering such a thing himself some day. His other research interests are in combinatorics and q-series.
Bates College, Lewiston, ME 04240
wjohnson@bates.edu