The Fundamental Theorem on Symmetric Polynomials: History’s First Whiff of Galois Theory

Ben Blum-Smith and Samuel Coskey

Ben Blum-Smith (ben@blumsmithmath.com) is a longtime math educator and current Ph.D. candidate in invariant theory at the Courant Institute of Mathematical Sciences at New York University. He leads professional development workshops for teachers through Math for America and blogs about math education at researchinpractice.wordpress.com. Blum-Smith is very excited that season two of Jane the Virgin is now on Netflix.

Samuel Coskey (scoskey@nylogic.org) is an assistant professor at Boise State University. He received his Ph.D. in set theory from Rutgers University in 2008. Coskey has taken part in many mathematical experiences for the general public and currently co-directs the Boise Math Circle. In his spare time he enjoys listening to and learning to play Tom Lehrer songs.

Evariste Galois’s (1811–1832) short life is one of the classic romantic tragedies of mathematical history. The teenage Galois developed a revolutionary theory of equations, answering more fully than ever before a centuries-old question: Why can we not find a formula for solving quintic polynomials analogous to the quadratic, cubic and quartic formulas? Then he died in a duel before his twenty-first birthday. His discoveries lay in obscurity for 14 years until Joseph Liouville encountered them, recognized their importance, and made them known [2, 10]. Over the next few decades, the ideas Galois introduced—groups and fields—transcended the problem they were designed to solve and reshaped the landscape of modern mathematics.

This story is told and retold in popularizations of mathematics. Less frequently discussed is the actual content of Galois’s discoveries. This is usually reserved for a course in advanced undergraduate or graduate algebra. This article is intended to give the reader a little of the flavor of Galois’s work through a theorem that plays a unique role in it. This theorem appears to have been understood, or at least intuited and used, by Newton, as early as 1665. By the turn of the nineteenth century it was regarded as well known. For Galois himself, it was the essential lemma on which his entire theory rested. This theorem is now known as the fundamental theorem on symmetric polynomials (FTSP).

This essay has three goals: the first expository, the second pedagogical, and the third mathematical. Our expository goal is to articulate the central insight of Galois theory—the connection between symmetry and expressibility, described below—by examining the FTSP and its proof. Here we intend to reach any mathematics students or interested laypeople who have heard about this mysterious “Galois theory” and wish to know what it is about. Our point of view (elaborated in the next and final sections)
is that the FTSP manifests the central insight of the theory, so that the interested reader can get a little taste of Galois theory from this one theorem alone.

We also wish to reach readers who have studied Galois theory but feel they missed the forest for the trees. After all, Galois theory has been substantially reformulated since Galois’s time. For example, Galois’s reliance on the FTSP has been replaced with the elementary theory of vector spaces over a field, a theory unavailable in the 1820s. A student of the modern theory may not even immediately recognize what we are calling the central insight—the connection between symmetry and expressibility—in what they have learned. In the final section we address this by placing the FTSP in the context of the theorems Galois proved using it and, in turn, link these to the modern formulation.

Our pedagogical aim comes from the approach we take to the theorem. Our narrative arose out of an informal inquiry-based course in group theory and the historical foundations of Galois theory. In it, we posed the problem of trying to give a naïve proof of the theorem before learning the classical proof. In the next two sections we describe the participants’ encounter with this problem and hope to showcase the pleasure of mathematical discovery as well as provide a classroom module for other instructors and students.

Our mathematical goals arise directly from this pedagogical experience. The classical proof of the FTSP, which we present below, involves a clever trick that diverges from the participants’ proof ideas and is therefore, from a pedagogical standpoint, a bit of a *deus ex machina*. The participants’ work in the course inspired us to develop a new proof that replaces this trick with another method that is more consonant with the direction of the participants’ thinking. Our view is that the new arguments shed light on what the classical proof was really doing all along. The explication of these proof variants and their relationship to the classical proof is our mathematical aim.

The back story

A symmetric polynomial in \( n \) variables \( x_1, \ldots, x_n \) is one that remains the same no matter how the variables are permuted. Some particularly simple symmetric polynomials are the \( n \) elementary symmetric polynomials:

\[
\begin{align*}
\sigma_1 &= x_1 + \cdots + x_n, \\
\sigma_2 &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n = \sum_{i<j} x_i x_j, \\
\sigma_3 &= \sum_{i<j<k} x_i x_j x_k, \quad \ldots, \quad \sigma_n &= x_1 x_2 \cdots x_n.
\end{align*}
\]

The FTSP states that every symmetric polynomial can be expressed uniquely in terms of these.

**Theorem 1 (Fundamental theorem on symmetric polynomials).** Any symmetric polynomial in \( n \) variables \( x_1, \ldots, x_n \) is representable in a unique way as a polynomial in the elementary symmetric polynomials \( \sigma_1, \ldots, \sigma_n \).

For example, since the polynomial \( d = (x_1 - x_2)^2 \) is unchanged by transposing the two variables, the theorem guarantees \( d \) can be expressed in terms of \( \sigma_1 = x_1 + x_2 \) and \( \sigma_2 = x_1 x_2 \). In this case the expression is easy to find: \( d = (x_1 + x_2)^2 - 4x_1 x_2 = \sigma_1^2 - 4 \sigma_2 \).

The importance of the theorem to the theory of equations stems from the fact known as Vieta’s theorem which shows that the coefficients of a single-variable polynomial are precisely the elementary symmetric polynomials in its roots.
Theorem 2 (Vieta’s theorem). Let \( p(z) \) be an \( n \)th degree monic polynomial with roots \( \alpha_1, \alpha_2, \ldots, \alpha_n \). Let \( \sigma_1, \ldots, \sigma_n \) be the \( n \) elementary symmetric polynomials in the \( \alpha_i \). Then

\[
p(z) = z^n - \sigma_1 z^{n-1} + \sigma_2 z^{n-2} - \cdots + (-1)^n \sigma_n.
\]

The proof is a straightforward computation, but its ease belies its significance. With this fact in hand, the FTSP becomes the statement that given any polynomial equation \( p(z) = 0 \), any symmetric polynomial in its roots is actually a polynomial in its coefficients, which can be written down without (in fact, on the way to) solving the equation. Continuing the example above, if \( x_1 \) and \( x_2 \) are the roots of a monic quadratic polynomial, then that polynomial is \( p(z) = z^2 - \sigma_1 z + \sigma_2 \). The theorem guarantees that the discriminant \( d \) of \( p(z) \) (defined as the square of the difference between the roots) would have an expression in terms of the coefficients. This of course is key to the quadratic’s solution: \( \sqrt{d} \) is the difference between the roots and \( \sigma_1 \) is the sum of the roots, and the roots themselves can be deduced from these two values. Since \( d \) can be expressed in terms of the coefficients, it follows that the roots can too.

This is the form in which the FTSP played its seminal historical role. It appears to have been at least intuited by Newton [2, pp. 6–8] who also gave a formula (now known as Newton’s theorem) that effectively proves FTSP for the special case of power sums \( p_j = \sum x_i^j \). The result embedded itself in the common knowledge of mathematicians over the course of the eighteenth century, through the calculations of many different people [2, 12]. For a discussion of some of its historical applications prior to Galois’s work, see [9].

The FTSP brings out one of the central insights of Galois theory, the connection between symmetry and rational expressibility. We have a polynomial \( p(z) \) whose coefficients we know. Even if we do not know the roots, the FTSP tells us that symmetric expressions in the roots are rationally expressible in terms of the coefficients. As a corollary, if the coefficients of \( p(z) \) are rational numbers, then every symmetric expression in the roots (e.g., the sum of their squares) has a rational value as well. Symmetry guarantees rational expressibility. In the last section we will indicate how this fits into the bigger picture of Galois theory.

In our course on Galois theory, we did not approach the FTSP directly, but rather sidled up to it by considering some problems of historical significance that implicitly depend on it. The first was a problem of Newton: Given two polynomials \( f, g \), how can one determine whether they have a root in common without finding the roots? (This problem is discussed at length in [9].) The second was posed by Gauss: Given a polynomial \( f \), without finding its roots, determine a polynomial \( g \) whose roots are the squares, or cubes, etc., of the roots of \( f \).

Participants solved both of these problems for polynomials of low degree. They wrote the desired expressions in the roots, which turn out to be symmetric, then expressed these in terms of the coefficients instead. For example, they considered Gauss’s problem for the quadratic \( f(z) = z^2 - \sigma_1 z + \sigma_2 \). How can we find \( g \) whose roots are the squares of the roots of \( f \)? Writing \( \alpha_1, \alpha_2 \) for the roots of \( f \), we need

\[
g = (z - \alpha_1^2)(z - \alpha_2^2) = z^2 - (\alpha_1^2 + \alpha_2^2)z + \alpha_1^2\alpha_2^2.
\]

To write down this polynomial without actually solving \( f \), we need expressions for the coefficients \( \alpha_1^2 + \alpha_2^2 \) and \( \alpha_1^2\alpha_2^2 \) in terms of \( \sigma_1, \sigma_2 \), the coefficients of \( f \). You may enjoy looking for them yourself before reading the next line.
\[ \alpha_1^2 + \alpha_2^2 = \sigma_1^2 - 2\sigma_2, \quad \alpha_1^2\alpha_2^2 = \sigma_2^2. \]

Participants were able to find such expressions in every case we considered and so began to suspect that something like the FTSP would be true. It was clear that any expression in the roots of a polynomial would have to be symmetric to be expressible in terms of the coefficients, since the coefficients are already symmetric. But it was not clear that any symmetric expression in the roots would be expressible in the coefficients.

The two and three variable cases

In this section we begin to approach the question of why any symmetric expression in the roots is expressible in terms of the coefficients. It is natural to begin with the special cases in which the polynomial has just two and then three variables. The participants were able to cobble together proofs in these two cases over the course of two meetings.

To start, let \( p(x, y) \) be a polynomial which is symmetric in \( x \) and \( y \). We want to show that it can be expressed as a polynomial in \( \sigma_1 = x + y \) and \( \sigma_2 = xy \). Taking an arbitrary monomial \( x^m y^n \) which appears in \( p(x, y) \), we will “take care of it” by expressing it in terms of \( \sigma_1 \) and \( \sigma_2 \). Renaming the variables if necessary, we can suppose that \( m \geq n \). If \( n > 0 \), then we can already write \( x^m y^n \) as \( \sigma_2^n x^{m-n} \), so it suffices to deal with monomials of the form \( x^n \). Note that the symmetry of \( p(x, y) \) implies its conjugate monomial \( y^n \) is also a term of \( p(x, y) \), so we can deal with \( x^n + y^n \) together. Now, we recognize \( x^n + y^n \) as the first and last terms of \( \sigma_1^n = (x + y)^n \). Hence, we have

\[
 x^n + y^n = \sigma_1^n - \binom{n}{1} x y^{n-1} - \cdots - \binom{n}{n-1} x^{n-1} y = \sigma_1^n - \sigma_2 q(x, y)
\]

where \( q(x, y) \) is a polynomial of degree \( n-2 \). This shows that an induction on the degree of \( p(x, y) \) will succeed.

In the case of three variables, let \( p(x, y, z) \) be a polynomial which is symmetric in \( x, y, z \). We wish to express it as a function of \( \sigma_1 = x + y + z \), \( \sigma_2 = xy + xz + yz \), and \( \sigma_3 = xyz \). Again consider an arbitrary monomial \( x^m y^n z^p \) in \( p(x, y, z) \) where for convenience we assume that \( m \geq n \geq p \). If \( p > 0 \), then we can write \( x^m y^n z^p \) as \( \sigma_3^p x^{m-p} y^{n-p} \), leaving a monomial with just two variables to deal with. In other words, we only need to treat monomials of the form \( x^m y^n \). Now, all of the conjugate monomials \( x^n z^m, x^m z^n, x^{m} z^{m}, y^{m} z^{n}, \) and \( y^{n} z^{m} \) are also found in \( p(x, y, z) \). In analogy to the two variable case, we now recognize that these are all terms of \( \sigma_1^{m-n} \sigma_2^n = (x + y + z)^{m-n} (xy + xz + yz)^n \). Thus, we can write

\[
 x^m y^n + x^n z^m + x^m z^n + x^n z^m + y^m z^n + y^n z^m = \sigma_1^{m-n} \sigma_2^n - q(x, y, z).
\]

Unlike the two variable case, the “leftover” terms \( q(x, y, z) \) need not have a common factor. However, any term of \( q(x, y, z) \) which happens to involve just two variables must be a conjugate of \( x^k y^\ell \) where \( m > k \geq \ell > n \) and \( k + \ell = m + n \). So while we have not reduced the degree in every case, in the cases where we have not we have nonetheless improved the situation in one key way: We have reduced the spread between the exponents. In other words, this time we will succeed using an induction which takes into account both the degree and the spread between the exponents in the case of monomials with just two variables.

It is natural to try to generalize this method to four and more variables, but there are some difficulties. For starters, it is not clear what the “spread between the exponents”
would mean when there are more than two variables in play! As in the class, we now present one of the standard proofs of FSTP, but will come back to this idea.

A classical proof

Our presentation follows [11]; the proof explicated here goes back at least to an 1816 paper of Gauss [4, paragraphs 3–5] with some key ideas tracing back to Waring in 1770 [12, p. 99].

Proof of the FTSP. Let \( f \) be the symmetric polynomial to be represented. We can assume without loss of generality that \( f \) is homogeneous, i.e., that all its terms have the same degree: This is because if \( f \) is symmetric, then the sum of terms \( f_d \) of \( f \) of a given degree \( d \) itself forms a symmetric polynomial. We can therefore represent each of the \( f_d \) in the \( \sigma_i \) individually.

Now, order the terms of \( f \) lexicographically. That is, put the term with the highest power of \( x_1 \) first and, in case of a tie, decide in favor of the term with the most \( x_2 \), and so on. Formally, define \( ax_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} > bx_1^{j_1}x_2^{j_2} \cdots x_n^{j_n} \) if \( i_1 > j_1 \), or if \( i_1 = j_1 \) and \( i_2 > j_2 \), or if \( i_1 = j_1 \), \( i_2 = j_2 \) and \( i_3 > j_3 \), etc., and then order the terms of \( f \) so that the first term is lexicographically greater than the second which is lexicographically greater than the third, and so on.

Because \( f \) is symmetric, for every term \( cx_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} \) in it, it also contains all possible terms that look like this one except with the variables permuted (its conjugates). It follows that the leading term of \( f \), say \( c_1x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} \), has \( i_1 \geq i_2 \geq \cdots \geq i_n \). Let

\[
g_1 = c_1\sigma_1^{i_1-i_2} \sigma_2^{i_2-i_3} \cdots \sigma_{n-1}^{i_{n-1}-i_n} \sigma_n^{i_n}
\]

which is symmetric and has the same leading term as \( f \). Thus \( f - g_1 \) is symmetric with a lower leading term which we denote \( c_2x_1^{j_1}x_2^{j_2} \cdots x_n^{j_n} \). As before, it follows from the symmetry that \( j_1 \geq j_2 \geq \cdots \geq j_n \). Thus we can let \( g_2 = c_2\sigma_1^{j_1-j_2} \sigma_2^{j_2-j_3} \cdots \sigma_n^{j_n} \) so that \( g_2 \) has the same leading term as \( f - g_1 \), and \( f - g_1 - g_2 \) has a leading term that is lower still.

Continue in like manner. The algorithm must eventually terminate with no terms remaining as there are only finitely many possible monomials \( x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} \) of a given degree. Thus we must come to a point where we have \( f - g_1 - g_2 - \cdots - g_k = 0 \), so \( f = g_1 + g_2 + \cdots + g_k \) is the desired representation of \( f \) as a polynomial in the \( \sigma_i \).

To prove uniqueness, it suffices to show that the zero polynomial in \( x_1, \ldots, x_n \) is representable uniquely as the zero polynomial in \( \sigma_1, \ldots, \sigma_n \). This is so because no two distinct products of elementary symmetric polynomials \( \sigma_1^{k_1} \cdots \sigma_n^{k_n} \) have the same leading term. (The leading term of \( \sigma_1^{k_1} \cdots \sigma_n^{k_n} \) is \( x_1^{k_1} \cdots x_n^{k_n} \) and the map \( (k_1, \ldots, k_n) \mapsto (k_1 + \cdots + k_n, \ldots, k_{n-1} + k_n, k_n) \) is injective.) Thus the leading terms in a sum of distinct products of elementary symmetric polynomials cannot cancel and such a sum cannot equal zero unless it is empty.

This lexicographic order argument is elegant, simple, and highly constructive. From a pedagogical standpoint, however, it depends on a very counterintuitive move. The lexicographic order (lex order for short) is a total order on the set of monomials. (In fact, it is a monomial order, meaning it is a well-order that is compatible with multiplication.) It determines a unique leading term in any polynomial and this fact is (prima facie) part of how the proof works. The proof conjures in one’s mind an image of the
terms of $f$ totally ordered and then picked off one-by-one, left to right, by our careful choice of $g_1, \ldots, g_k$.

However, since $f$ and $g_1, \ldots, g_k$ are all symmetric, the terms are not really being picked off one at a time. Forming $f - g_1$ not only cancels the leading term $c_1x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}$ but all of its conjugates as well (for instance, the “trailing term” $c_1x_1^{i_n}x_2^{i_{n-1}} \cdots x_n^{i_1}$). Somehow, the lex ordering obscures the symmetry between the conjugates by distinguishing one of them as the leading term, even while it exploits this symmetry to make the proof work.

In this way it diverges sharply from the participants’ naïve attempts to prove the theorem, all of which dealt with all the monomials in a given conjugacy class on equal footing. This makes the appeal to lex order highly unexpected, which is part of the proof’s charm, but it also raises the (essentially mathematical, but pedagogically resonant) question of whether it is possible to give a version of the proof without this unexpected disruption of symmetry.

To look at it from another angle, the order in which the algorithm given in this proof operates on the terms of $f$ is not actually the lex order. Rather it is the order that lex order induces on the set of conjugacy classes of terms. That is, the first conjugacy class is the one containing the lexicographically leading term, the second contains the lexicographically highest-ranking term not contained in the first, etc. We could call this symmetric lexicographic order. Note that it is no longer a total order on the monomials (only on the conjugacy classes). Thus the proof’s appeal to lex order is somehow deceptive. The real order is something else. From this angle, the pedagogical question becomes, are there descriptions of symmetric lex order that do not pass through actual lex order?

We note that many proofs of the FTSP are known and they do not all share the surprising symmetry-breaking feel of the lex proof; some of our favorites are given in [1, 2, 7]. In fact, one can derive the FTSP from Galois theory itself, rather than the reverse, because the modern development of the latter no longer depends on the former, as in [6]. The lex proof nonetheless stands out as especially constructive, in that the algorithm it gives is practical for writing symmetric polynomials in terms of the elementary ones; short; and enduringly popular.

The dissonance between the participants’ approach and the one taken in this classical proof led us to return to the idea of “spread between the exponents” mentioned in the last section. This idea ultimately brought answers to the above questions, in the form of both an alternative proof and a much richer understanding of the classical proof.

**Spreadness**

We return to the ideas of our proof in the two and three variable cases and develop them into a complete argument. Recall that to generalize our ideas, we first need to overcome the difficulty of deciding what the spread between the exponents means when there is a larger number of variables. Indeed, finding this definition is the linchpin of our strategy. We will give a definition (and later, a family of definitions) that allow us to prove the theorem by building an algorithm that picks off the monomials with the most spread-out exponents first. The algorithm is identical in spirit and similar in practice to the standard one, but uses spread-out-ness (what we henceforth call spreads) rather than lex order to determine which monomials to cancel out first. It thus carries out the classical proof’s program while avoiding the symmetry disruption imposed by the lexicographic ordering (answering “yes” to our first pedagogically resonant question above).
In our proof of the two and three variable case, our initial definition of spread was the highest exponent minus lowest. Unfortunately, a simple computation shows this will not work in general. In terms of statistics, it is analogous to the range of the dataset of exponents of a given monomial $x_1^{i_1} \cdots x_n^{i_n}$. But the range is not a good measure of dispersion because it does not involve all of the exponents. Instead we consider the following.

**Definition.** The *spreadness* of a monomial $x_1^{i_1} \cdots x_n^{i_n}$ is the sum $i_1^2 + \cdots + i_n^2$.

Again in terms of statistics, this is equivalent to the variance of the dataset of exponents (in that it induces the same ordering; see the end of this section for an elaboration). The spreadness is also equivalent to the height of the center of gravity of the monomial when it is pictured as a pile of bricks with a stack of $i_k$ bricks corresponding to each $x_k$ (see Figure 1). Moreover, it is a nonnegative integer, allowing us to use it as the basis of an induction argument.

The key fact to establish is that just as $c_1 x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ with $i_1 \geq i_2 \geq \cdots \geq i_n$ is the leading term of $c_1 \sigma_1^{i_1-i_2} \sigma_2^{i_2-i_3} \cdots \sigma_n^{i_n}$ when the terms are ordered lexicographically, it and all its conjugates also have strictly greater spreadness than the rest of the terms of this latter product.

**Theorem 3 (Spreadness lemma).** Given $i_1, \ldots, i_n$ with $i_1 \geq \cdots \geq i_n$, the terms of $\sigma_1^{i_1-i_2} \sigma_2^{i_2-i_3} \cdots \sigma_n^{i_n}$ with maximum spreadness are precisely $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ and its conjugates.

*Proof.* In this argument we identify a monomial $x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}$ with a sequence of stacks of heights $j_1, \ldots, j_n$ of identical bricks. We first compute that for terms taken from $\sigma_1^{i_1-i_2} \sigma_2^{i_2-i_3} \cdots \sigma_n^{i_n}$, the spreadness is an increasing linear function of the vertical coordinate of the center of gravity of its corresponding brick configuration. Supposing that each brick has unit mass, then the vertical coordinate of the center of gravity is given by the sum over the bricks of each brick’s height divided by the number of bricks. If we suppose the first brick of each stack lies at height 1 and each brick has unit height, then the stack of height $j_1$ contributes $1 + \cdots + j_1 = j_1 (j_1 + 1)/2$ to the sum. The vertical coordinate $y$ of the center of gravity is then given by

$$y = \frac{1}{d} \left( \frac{j_1 (j_1 + 1)}{2} + \cdots + \frac{j_n (j_n + 1)}{2} \right)$$

$$= \frac{1}{2d} \left( j_1^2 + \cdots + j_n^2 + j_1 + \cdots + j_n \right)$$

$$= \frac{1}{2d} (s + d)$$

where $d$ is the number of bricks (i.e., the degree of the monomial) and $s$ is the spreadness. So $s = 2dy - d$ and, since $d$ is fixed, $s$ is an increasing linear function of $y$ as claimed.

Next, we observe that all of the terms of $\sigma_1^{i_1-i_2} \sigma_2^{i_2-i_3} \cdots \sigma_n^{i_n}$ can be obtained from $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ by moving some bricks horizontally (and dropping them onto the top of the stack below if necessary). The conjugates of $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ are precisely those terms in which each layer of bricks rests completely on top of the layer below it before any dropping takes place. Thus bricks will fall for precisely those terms that are not conjugates of $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$; see Figure 1.
Finally, we appeal to the simple fact that given any physical configuration of bricks, moving some bricks to lower positions decreases the center of gravity.

Once this is established, the proof of the fundamental theorem follows the outline of the standard argument given above.

Proof of the FTSP using the spreadness lemma. Let \( f \) be the symmetric polynomial to be represented. As above, we lose nothing by assuming \( f \) is homogeneous.

The algorithm proceeds as in the standard proof except with spreadness playing the role of lexicographic order. Pick any term of \( f \) with maximum spreadness \( s_1 \) and consider it and its conjugates. Form the product of elementary symmetric polynomials \( g_1 \) that has these terms as its terms of maximum spreadness. (If the terms of \( f \) have coefficients \( c_i \) and exponents \( i_1 \geq i_2 \geq \cdots \geq i_n \), then \( g_1 = c_1 \sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_n^{i_n} \) as always.) Then since these terms are the only terms of \( g_1 \) with spreadness as high as \( s_1 \), by the spreadness lemma \( f - g_1 \) contains fewer terms of spreadness \( s_1 \) than \( f \) does, possibly zero.

Continuing in like manner with \( f - g_1 \), forming \( g_2 \), and then \( f - g_1 - g_2 \), etc., we get an algorithm that must terminate because at each stage, either the maximum spreadness or the number of terms with this spreadness has been decreased.

The uniqueness of the representation follows exactly as in the standard proof.

To connect spreadness and variance, we compute that for terms taken from \( \sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_n^{i_n} \), the spreadness \( s \) is an increasing linear function of the variance \( \sigma^2 \) of the set \( \{ j_1, j_2, \ldots, j_n \} \). Indeed,

\[
s = n\sigma^2 + n\mu^2 \]

Here \( n \) is fixed and so is the mean \( \mu \), being a function of just \( n \) and the degree \( d \). Thus, \( s = n\sigma^2 + n\mu^2 \) is an increasing linear function of \( \sigma^2 \).

The spreadness and lex orderings

It is natural to ask whether there is any relationship between the spreadness and lexicographic orderings on monomials. Apropos of our discussion after the classical proof, the more natural comparison is between spreadness and what we there defined as the symmetric lexicographic order, i.e., the order that lex induces on conjugacy classes of monomials. In the spreadness lemma, we have shown that the two orderings single out the same conjugacy class of monomials as leading among those that occur in a single product of the form \( \sigma_1^{i_1 - i_2} \sigma_2^{i_2 - i_3} \cdots \sigma_n^{i_n} \).

In general, however, the two orderings do not agree. For example, \( x_1^3x_2x_3x_4x_5x_6 \) beats \( x_1^2x_2^2x_3x_4^3 \) lexicographically, but has a lower spreadness, 14 versus 16.
Still, this can be remedied by replacing the spreadness with the _\textit{pth moment spreadness}_ (that is, \(i_1^p + \cdots + i_n^p\)) for suitably large \(p\). In the above examples, letting \(p = 3\), the new score becomes 86 versus 64. We show next that this can be done generally.

**Theorem 4.** Symmetric lex order is the limit of the order on conjugacy classes of monomials given by the _\textit{pth moment spreadness}_ as \(p \to \infty\), in the sense that given any finite set of classes, for all sufficiently high \(p\) the _\textit{pth moment order}_ on those classes matches the symmetric lex order.

**Proof.** Let \(x_1^i \cdots x_n^i\) and \(x_1^j \cdots x_n^j\) be given with \(i_1 \geq \cdots \geq i_n\) and \(j_1 \geq \cdots \geq j_n\). Assume that \(x_1^i \cdots x_n^i\) precedes \(x_1^j \cdots x_n^j\) lexicographically and let \(k\) be minimal with \(i_k > j_k\). Then we may choose \(p\) large enough that \(i_k^p > nj_k^p\) and it follows easily that \(i_1^p + \cdots + i_n^p > j_1^p + \cdots + j_n^p\).

This satisfies our second mathematical-but-pedagogically-motivated question from after the classical proof: a way to characterize symmetric lex order without passing through lex order. To our taste, this characterization shows that symmetric lex order is “more natural” than is obvious from its definition via (actual) lex order.

Moreover it is possible to give a version of the spreadness lemma for any of the higher moments, although the proof is somewhat more involved without the center-of-gravity interpretation available.

**Theorem 5 (Spreadness lemma for higher moments).** Given \(i_1, \ldots, i_n\) with \(i_1 \geq \cdots \geq i_n\), the terms of \(\sigma_{i_1}^{1-i_2} \sigma_{i_2}^{1-i_3} \cdots \sigma_{i_n}^{1}\) with maximum _\textit{pth moment spreadness},_ for \(p > 1\), are precisely \(x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}\) and its conjugates.

**Proof outline.** The terms \(x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}\) of \(\sigma_{i_1}^{1-i_2} \sigma_{i_2}^{1-i_3} \cdots \sigma_{i_n}^{1}\) all satisfy the following conditions: every exponent \(j_k\) is at most \(i_k\), every sum of two exponents \(j_k + j_{\ell}\) is at most \(i_k + i_\ell\), and more generally every sum of \(\ell\) exponents is at most \(i_1 + \cdots + i_\ell\), with equality when \(\ell = n\). Thus each term corresponds to a lattice point \((j_1, j_2, \ldots, j_n)\) in the first quadrant of \(\mathbb{R}^n\), contained in the convex polytope \(P\) cut out by the inequalities

\[
z_{k_1} + \cdots + z_{k_\ell} \leq i_1 + \cdots + i_\ell
\]

for all \(\ell\) and all sequences \(k_1 < \cdots < k_\ell \leq n\). Furthermore, the term \(x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}\) and its conjugates correspond exactly to those lattice points that realize equality in each of the above inequalities for some choice of \(k_i\). In other words, they correspond to the vertices of the convex polytope \(P\). This is because, in the language and imagery of the center of gravity proof of the spreadness lemma, equality is realized in each inequality (for a maximizing choice of \(k_i\)) if and only if no brick has fallen. If a brick in the \(\ell\)th highest stack falls to a lower stack, this implies that the highest \(\ell\) stacks now have a lower total than they did originally.

Now we appeal to the fact that the _\textit{pth moment spreadness}_ is a monotone function of the \(L^p\) norm on \(\mathbb{R}^n\) and is strictly subadditive for \(p > 1\), i.e., equality holds in \(\|u + v\| \leq \|u\| + \|v\|\) only when one of \(u, v\) is a nonnegative multiple of the other. It follows that if \(\|u\| = \|v\|\) and \(u \neq v\), then any nontrivial convex combination of \(u, v\) has strictly smaller norm than \(u, v\) have. (One sees this by applying the inequality to \(\|\mu u + \nu v\|\) with \(\mu, \nu > 0\) and \(\mu + \nu = 1\).) More generally, if the extreme points of a convex body all have the same norm, all the other points of the body must have strictly smaller norm. In the present case, the tuples \((j_1, \ldots, j_n)\) corresponding to \(x_1^{j_1} \cdots x_n^{j_n}\) and its conjugates all have the same \(L^p\) norm \(\sqrt[p]{i_1^p + \cdots + i_n^p}\). Since they are the vertices of a convex polytope containing the tuples corresponding to all the
other terms, these other tuples must have smaller $L^p$ norm and therefore smaller $p$th moment spreadness.

Thus, the FTSP can be proven using the $p$th moment spreadness for any $p > 1$. The spreadness proof given previously is only the first in an infinite sequence of nearly identical proofs, and the classical proof is, by Theorem 4, in some sense the last. All the proofs have in common an algorithm that represents an arbitrary symmetric polynomial $f$ by forming products of elementary symmetric polynomials $\sigma_k$ in a way that mimics the terms of $f$ with maximum exponent dispersion first. Thus they are all fundamentally inductions on the extent of exponent dispersion—hence "spreadness." Each proof measures exponent dispersion a little differently but they all agree about the terms of maximum dispersion in expansions of monomials in the $\sigma_k$. They all agree because these terms correspond to the extreme points of certain convex polytopes in $\mathbb{R}^n$, although we have other, easier ways to see this in the special first and last cases. Since the order in which the classical algorithm operates on $f$ comes from the limit of these ways of measuring, we can see it in some sense as having measured exponent dispersion all along!

The FTSP in Galois’s work

In this concluding section we place the FTSP in the greater context of Galois theory by showing how it is an example of a larger phenomenon. The FTSP is a statement about polynomials, but it is easy to extend it to all rational expressions [1, p. 551]. With this extension, the FTSP says that expressions that are completely symmetric are completely rationally expressible. In his seminal essay Mémoire sur les conditions de résolubilité des équations par radicaux, Galois developed a chain of results that tie types of partial symmetry to types of partial rational expressibility as well. We give them without proof. He revealed the FTSP as just the first link in a chain of statements that tie types of symmetry to forms of rational expressibility. The second chain link (like the FTSP itself) was already well known in Galois’s time.

**Theorem 6.** If $f$ is a rational function of $x_1, \ldots, x_n$ that is symmetric under all permutations of the $x_i$ that fix $x_1$, then it is expressible as a rational function of $\sigma_1, \ldots, \sigma_n$ and $x_1$.

The third appears as Lemma III in Galois’s essay. It is a consequence of a 1771 theorem of Lagrange [2, pp. 32–37] & [13, pp. 80–81], but Galois’s argument is independent of Lagrange’s [2, pp. 43–5].

**Theorem 7.** If $V$ is a rational function of $x_1, \ldots, x_n$ that is not fixed by any nontrivial permutation of the $x_i$, then every rational function of the $x_i$ is expressible as a rational function of the $\sigma_i$ and $V$.

See Table 1 for a summary of these results.

The final link in the chain is Galois’s famous Proposition I. Theorems 1, 6, and 7 are all simultaneously lemmas for and special cases of this grand result, which forms one half of what is now called the fundamental theorem of Galois theory. The following paraphrases Galois’s statement.

**Galois’s Proposition I.** Let $f$ be a polynomial with coefficients $\sigma_1, \ldots, \sigma_n$. Let $x_1, \ldots, x_n$ be its roots. Let $U, V, \ldots$ be some other numbers that are rational functions of the $x_i$. Then, there exists a group $G$ of permutations of the $x_i$ such that the rational functions of the $x_i$ fixed under all the permutations in $G$ are exactly those whose values are rationally expressible in terms of $\sigma_1, \ldots, \sigma_n$ and $U, V, \ldots$.
Table 1. Symmetry and expressibility results for a rational function $f$ of $x_1, \ldots, x_n$.

| If it is invariant under . . . | then it is rationally expressible in . . . |
|-------------------------------|------------------------------------------|
| all permutations               | $\sigma_1, \ldots, \sigma_n$            |
| all permutations that fix $x_1$| $\sigma_1, \ldots, \sigma_n$, and $x_1$  |
| any subset, or no permutations at all | $\sigma_1, \ldots, \sigma_n$, and $V$   |

We think of the numbers $U, V, \ldots$ in the statement of the proposition as specifying the type of rational expressibility being allowed. Thus the proposition is stating that no matter what type of rational expressibility (choice of $U, V, \ldots$) you want to allow, there exists a type of symmetry (specified by the group $G$) that coincides perfectly with that type of expressibility.

Even if you have studied Galois theory, this formulation may be unfamiliar. The connection to what you have seen before is that by “type of rational expressibility” we really mean field. The set of quantities that are rational functions of the coefficients $\sigma_1, \ldots, \sigma_n$ forms the coefficient field of $f$: all the numbers you can write down rationally if you can write down $f$. Similarly the set of quantities that are rational functions of the roots $x_1, \ldots, x_n$ form the splitting field of $f$: everything you can write down if you can solve $f$. By allowing the numbers $U, V, \ldots$ along with the coefficients $\sigma_i$ in your rational expressions, you get some field that contains the coefficient field and lies inside the splitting field. So we can state Galois’s Proposition I in the following modern way: Given a polynomial $f$ and a field $K$ lying between the coefficient field and the splitting field of $f$, there exists a group $G$ of permutations of the roots whose action on the splitting field of $f$ has fixed field $K$.

We close with two remarks. First, the result of Proposition I is just half of what we now call the fundamental theorem of Galois theory. The other half states that if you find the group $G$ corresponding to the coefficient field itself (called the Galois group of $f$), then every subgroup of $G$ corresponds to some intermediate field $K$. There is thus a one-to-one correspondence between fields intermediate between the coefficients of $f$ and splitting fields, on the one hand, and subgroups of the Galois group of $f$, on the other.

Second, up until the statement of Proposition I, the $x_i$ have been formal symbols and the $\sigma_i$ have been formal polynomials in them, but for this statement the $\sigma_i$ are prior to the $x_i$ and may be elements of any field containing $\mathbb{Q}$. Galois tacitly assumes that the roots $x_i$ of $f$ exist, somewhere, in some sense. Today we would say he assumes the existence of a splitting field. Most mathematicians prior to the nineteenth century working in algebra made this same assumption without question. Gauss famously argued that it needed justification, in motivating his many proofs of the fundamental theorem of algebra, that every integer polynomial splits into linear and quadratic factors over $\mathbb{R}$. In fact, one of these proofs was the primary goal of the paper in which Gauss published the lexicographic order proof of the FTSP [4]! See [3, pp. 912–913] and [13, pp. 94–102].

In closing, we hope to have shown you that the FTSP contains the first whisper of Galois’s connection between symmetry and rational expressibility. If you are interested to learn more, Edwards [3] explicates some of Galois’s own proofs of the above propositions in modern language and deals with a number of details we have elided here for reasons of length. This article is perhaps best appreciated alongside Galois’s original essay, which is printed in English translation in several sources [2, 5, 8].
Acknowledgment. The course mentioned above was given by the first author in 2009–2010 to a small group of teachers and mathematicians including the second author, Kayty Himmelstein, Jesse Johnson, Justin Lanier, and Anna Weltman. We are grateful for their active participation. We would also like to thank Benjamin Weiss for his assistance tracking the history of the FTSP, Harold Edwards for a clarifying conversation about Galois’s Proposition I, Walter Stromquist for very helpful comments including the insight behind Theorem 4, and several anonymous referees for very helpful comments.

Summary. We describe the fundamental theorem on symmetric polynomials (FTSP), exposit a classical proof, and offer a novel proof that arose out of an informal course on group theory. The paper develops this proof in tandem with the pedagogical context that led to it. We also discuss the role of the FTSP both as a lemma in the original historical development of Galois theory and as an early example of the connection between symmetry and expressibility that is described by the theory.

References
1. M. Artin, Algebra. Prentice Hall, Englewood Cliffs, NJ, 1991.
2. H. M. Edwards, Galois Theory. Springer, New York, 1984.
3. ———, Galois for 21st-century readers, Notices Amer. Math. Soc. 59 (2012) 912–923, http://dx.doi.org/10.1090/noti869.
4. C. F. Gauss, Demonstratio nova altera theorematis ommem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse, Comm. Recentiores 3 (1816) 107–142. Werke, vol. 3, 31–56.
5. S. Hawking, God Created the Integers: Mathematical Breakthroughs that Changed History. Running, Philadelphia, 2005.
6. T. W. Hungerford, Algebra. Eighth ed. Springer, New York, 2003.
7. S. Lang, Algebra. Second ed. Addison–Wesley, Reading, MA, 1984.
8. P. M. Neumann, The Mathematical Writings of Évariste Galois. European Mathematical Society, Zürich, 2011.
9. G. St. George, Symmetric polynomials in the work of Newton and Lagrange, Math. Mag. 76 (2003) 372–379.
10. J. Stillwell, Mathematics and its History. Third ed. Springer, New York, 2010.
11. B. Sturmfels, Algorithms in Invariant Theory. Second ed. Springer, Vienna, 2008.
12. J.-P. Tignol, Galois’ Theory of Algebraic Equations. World Scientific, Singapore, 2001.
13. B. L. van der Waerden, A History of Algebra: From Al-Khwarizmi to Emmy Noether. Springer, Berlin, 1985.