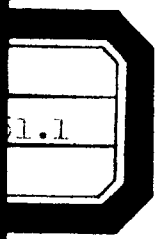
Graduate Texts in Mathematics 101

Harold M. Edwards

Galois Theory



Harold M. Edwards

Galois Theory

Springer-Verlag
New York Berlin Heidelberg Tokyo

Preface

This exposition of Galois theory was originally going to be Chapter 1 of the continuation of my book *Fermat's Last Theorem*, but it soon outgrew any reasonable bounds for an introductory chapter, and I decided to make it a separate book. However, this decision was prompted by more than just the length. Following the precepts of my sermon "Read the Masters!" [E2], I made the reading of Galois' original memoir a major part of my study of Galois theory, and I saw that the modern treatments of Galois theory lacked much of the simplicity and clarity of the original. Therefore I wanted to write about the theory in a way that would not only explain it, but explain it in terms close enough to Galois' own to make his memoir accessible to the reader, in the same way that I tried to make Riemann's memoir on the zeta function and Kummer's papers on Fermat's Last Theorem accessible in my earlier books, [E1] and [E3]. Clearly I could not do this within the confines of one expository chapter.

And so I decided to write a short book—a sort of volume 1½ of my work on Fermat's Last Theorem—devoted entirely to the basics of Galois theory. There is very little in this book that is not already to be found, however concisely and however lacking in proof, in Galois. The one major exception is the material on factorization of polynomials (§§49–61), which is due to Kronecker and which seems to me to be necessary to give clear meaning to the computations with roots of algebraic equations that Galois and Lagrange performed without inhibition and without comment.

The crux of Galois theory is, appropriately enough, Galois' Proposition I, which is the following characterization of what we call the *Galois group* of an equation. Let $a, b, c, \ldots$ be the n roots (assumed distinct) of an algebraic equation $f(x) = 0$ of degree n. The Galois group is a certain subgroup of the group of permutations of the roots $a, b, c, \ldots$. Galois used it to deter-

mine whether a given polynomial in the roots $F(a, b, c, \ldots)$ has a known value—in modern parlance, to determine whether $F(a, b, c, \ldots)$ is in the ground field. The characteristic property of the Galois group is that $F(a, b, c, \ldots)$ has a known value if and only if

$$F(a, b, c, \ldots) = F(Sa, Sb, Sc, \ldots)$$

for all permutations S of the Galois group. Galois proved the existence and uniqueness of a group with this property by *constructing* it, using what later became known as a Galois resolvent. (This characterization of the Galois group will be more recognizable to readers familiar with modern formulations of Galois theory after they read the first corollary in §41. See also §63.)

The major theorems of Galois, such as the theorems on the solvability of equations by radicals, flow from the study of the relationship between algebraic equations $f(x) = 0$ and the groups associated with them. Of particular importance is the analysis of the way in which the group is reduced when the field of known quantities is extended (Galois' Propositions II–IV).

Some recent texts on Galois theory place mistaken emphasis on the question of finding explicit quintic equations, with rational coefficients, which cannot be solved by radicals. This is a moderately interesting result (one not covered in this book) but it is not a key theorem of Galois theory. Galois showed that an algebraic equation is solvable by radicals if and only if the associated group is solvable. A given quintic with rational coefficients can therefore be tested for solvability. Abel's theorem that the *general* quintic is not solvable states that the equation $x^5 + Bx^4 + Cx^3 + Dx^2 + Ex + F = 0$—an equation with coefficients in the field $\mathbb{Q}(B, C, D, E, F)$ obtained by adjoining five transcendental elements (variables) to $\mathbb{Q}$—is not solvable by radicals. (In Galois theory this follows from the fact that the Galois group of this equation is the full group of 120 permutations of the five roots.) In other words, no field extension of $\mathbb{Q}(B, C, D, E, F)$ obtained by a succession of adjunctions of radicals can ever contain a root of the given equation. This is what it means to say that the quadratic formula

$$x = \frac{-B \pm \sqrt{B^2 - 4C}}{2},$$

and the much more complicated formulas for the cubic and quartic equations (Exercises 1 and 2 of the Sixth Set) have no generalization to the quintic equation.

Having just mentioned the exercises, I hasten to reassure the reader that *the exercises are not essential to the book*. The only proofs that are relegated to the exercises are those that I believe to be too easy, or too much like other proofs already covered, to spend time on in the text. Naturally, the reader who does the exercises will have a far greater understanding of the subject, and will learn many things not contained in the text, but to do all the exercises will surely consume an enormous amount of time. The reader who has just

read the text will have covered all the propositions and methods of proof that I consider to be basic to Galois theory.

What preparation do I assume on the part of the reader? Because terminology changes so much from decade to decade and from field to field, I have tried to assume as little terminology as possible. (When I completed my undergraduate degree 25 years ago, I had had courses in advanced calculus, determinants and matrices, differential equations, measure theory, complex variables, etc., but I had never encountered the definition of a group or an abstract vector space.) However, I have assumed a certain degree of mathematical *experience* on the part of the reader, by which I mean experience in computation and mathematical reasoning. The main theorems of Galois theory state, in the last analysis, that certain computations with polynomials produce certain results. In most cases the computations are too long to do, and the *idea* of the computation is what counts, not any particular cases of it. The reader should have enough mathematical experience (and talent) to be able to conceive a general computation and its properties after having done a few simple examples.

The approach of the book is consistently *algebraic* and *constructive*. The fields considered are those obtained from the rational numbers by adjoining a finite number of algebraic and/or transcendental elements. (Fields with characteristic p are mentioned only in passing. Fields obtained by completion processes—the real and complex numbers, algebraic extensions of p-adic fields—are not considered at all.) *The constructive approach implies that theorems mean what they say.* For example, when a theorem says that an equation is solvable, the proof must give a procedure—however impractical—for constructing a splitting field by the adjunction of radicals. I believe that this approach is very much in tune with Galois' conception of the subject.

Liouville, in the "Avertissment" preceding his publication of Galois' works in 1846, writes of the "vivid pleasure" he enjoyed when he realized that Galois' methods were correct and that his theorems could be rigorously proved. I experienced what I imagine was a similar—if lesser—pleasure when I realized that two parts of Galois' memoir, which I at first thought were mistakes, are perfectly correct. These are the places where Galois later wrote "*On jugera*", in the case of the first, and "Something in this proof needs to be completed—I haven't the time" in the case of the second.

The "*On jugera*" passage is the one where Galois proves the crucial lemma stating that any rational function of the roots can be expressed as a rational function of the Galois resolvent. Poisson had called Galois' proof "insufficient" but pointed out that the lemma followed from a theorem of Lagrange. Galois, rather than elucidate his proof, laconically replied, "That remains to be seen" (freely translated). My opinion is in §37.

The famous statement "I haven't the time" occurs in a marginal note Galois made, probably on the night before the duel, with regard to the proof of his Proposition II, which he said needed to be "completed". Although his

proof appears wrong at first because he adjoins *one* root r of an equation and then uses *other* roots of the equation, and although Liouville [Gl, p. 492] found it necessary to circumvent Galois' proof entirely, I believe now that the proof given in §44 is very close to what Galois had in mind, and that the marginal note was merely prompted by the fact that he had *changed the statement of the Proposition*, and realized that the proof needed to be amended accordingly. (In fact, the Proposition, as stated, is false. The index of the subgroup need not be 1 or p when p is not prime—it must simply be a divisor of p.) A similar situation occurred with Proposition III, where Galois again changed the statement, making it more general, at the last minute, and had only time enough to say, "One will find the proof."

Finally, I hope it is superfluous to add that, while I have said above that most of what is in this book is already in Galois, the converse is far from true. The book contains a rather complete account of Galois' main memoir, "Mémoire sur les conditions de résolubilité des équations par radicaux" (Appendix 1 contains a translation of this memoir) but it does not make any claim to cover his other works. These contain, I am told, remarkable insights into a number of topics, including the theory of Abelian functions and finite simple groups. I return to my perennial refrain: Read the masters.

Acknowledgments

My greatest indebtedness is to Mr. James M. Vaughn, Jr., and the Vaughn Foundation Fund. This book is a direct result of their encouragement and support, for which I am deeply grateful. Work on the book was also supported by a Fellowship of the John Simon Guggenheim Memorial Foundation during 1981/82. I am grateful for both the financial support and the honor of a Guggenheim Fellowship. A large number of friends and colleagues have helped me by reading and commenting on early versions of the manuscript. Some major revisions prompted by their criticisms have not been seen by any of them, so it is even truer than usual that they are entitled to credit for many improvements in the book but free from blame for its faults. I would especially like to thank the following for their help: Jay Goldman, Mel Hausner, Susan Landau, Richard Pollack, Walter Purkert, Michael Rosen, Gabriel Stolzenberg, René Taton and his associates Pr. Ch. Houzel and M. Y. Hirano, William Y. Vélez, B. L. van der Waerden, and an anonymous reader for Springer-Verlag. Finally, my thanks to New York University and the Courant Institute for their overall support and assistance, including a sabbatical year 1980/81, the excellent library, and the expert word-processing of Connie Engle.

Contents

Galois

§1 Great mathematicians usually have undramatic lives, or, more precisely, the drama of their lives lies in their mathematics and cannot be appreciated by nonmathematicians. The great exception to this rule is Evariste Galois (1811–1832). Galois' life story—what we know of it—is like a romantic novel. Although he was making important mathematical discoveries when he was still in secondary school, he was denied admission to the Ecole Polytechnique, which was the premier institution of higher learning in mathematics at the time, and the mathematical establishment ignored, mislaid, lost, and failed to understand his treatises. Meanwhile, he was persecuted for his political activities and spent many months in jail as a political prisoner. At the age of 20 he was killed in a duel involving, in some mysterious way, honor and a woman. On the eve of the fatal duel he wrote a letter to a friend outlining his mathematical accomplishments and asking that the friend try to bring his work to the attention of the mathematical world. Against great odds, Galois' few supporters did finally, 14 years after his death, succeed in finding an audience for his work, and portions of his writings were published in 1846 by Joseph Liouville in his *Journal de Mathematiques*. After that, recognition of the great importance of his work came very quickly, and Galois began to be regarded, as he is today, as one of the great creative mathematicians of all time.

§2 The purpose of this book is to convey the mathematical drama of Galois' work, so there will be no more mention of his short, unhappy life,* but

* For biographical information see Dupuy [D1], Kiernan [K1], Rothman [R1].

one point needs to be made about its most dramatic feature, namely, the fact that Galois was able, at such an early age and without the benefit of any formal higher education, to make discoveries that would win him lasting fame. Surely many aspiring young mathematicians have been discouraged by Galois' story, saying to themselves something like, "Here I am already x years old, $x - 20$ years older (younger) than Galois was when he *died*, and, although I like math and have always done well at it, I would no more be able to make a great discovery than I would be able to swim the Atlantic." How was Galois able to do it? Was he blessed with some superhuman gift that put him in a class apart? I think not. Of course, talent is essential, and few are as talented as Galois. Still, talent alone is not enough. Galois had to reach the point where he knew enough and had enough techniques at his command to be able to move beyond what had been done before. The secret of how he was able to do this is contained, I believe, in a passage in Dupuy's biography of Galois [D1, p. 206]: "Elementary algebra books never satisfied Galois because he didn't find in them the stamp of the inventors; right from his first year of mathematics he turned to Lagrange."

Lagrange's *Réflexions sur la Résolution Algébrique des Equations* (1771) is the treatise of Lagrange most likely to have inspired the creation of Galois theory. It is an extraordinary work, written in a relaxed, discursive style that was rather common in the eighteenth century, but is virtually unknown in mathematical writing today. It discusses at length the central question of the time in the theory of algebraic equations, namely: What is the essence of the methods by which it is possible to solve equations of degrees 2, 3, and 4? Is it possible to extend these methods to equations of higher degree and, if not, why not? Lagrange gave an insightful answer to the first question, describing the solutions of equations of low degree in terms of a unified technique now known as the technique of the *Lagrange resolvent*.* His answer to the second question, on the other hand, is quite inconclusive. He shows that the technique does not apply in an obvious way to equations of degree 5 or higher, and he discusses some techniques—notably the technique of permuting the roots of an algebraic equation—which are relevant to the applications of Lagrange resolvents to equations of higher degree, but he gives no final answer. In short, it is a paper that gives the reader as much information about the problem as the author can provide and indicates the direction which the author feels further work should take. Viewed in this way, Lagrange's paper seems the perfect source of inspiration for a Galois.

Thus, in order to appreciate Galois' theory, it is natural first to review Lagrange's work. We will go much farther back than that—all the way to ancient Babylon—and then review a few other aspects of the development of algebra before discussing the main features of the work of Lagrange and then moving on to his successors, Gauss and Galois.

* A very similar technique was used a few months earlier by Vandermonde (see §15), but this was unknown to Lagrange.

Quadratic Equations 1700 B.C.

§3 Archeological research in the twentieth century has revealed the surprising fact that the peoples of Mesopotamia in the period around* 1700 B.C. had an advanced mathematical culture, including an excellent sexagesimal system of arithmetic and a knowledge of the Pythagorean theorem (a millennium before Pythagoras!). Of particular relevance to the theory of equations and Galois theory is the knowledge in this ancient culture of a method for the solution of quadratic equations.

According to Neugebauer [N1], the technique commonly used in the Babylonian texts to solve quadratic equations can be viewed as a reduction to a normal form, followed by a method for solving the normal form. The normal form was to *find two numbers given their sum and their product.* In modern algebraic notation, this can be stated: Given two numbers p and s, and given that $xy = p, x + y = s$, find x and y. The steps by which the Babylonians solved this problem are as follows:

1. Take half of s.
2. Square the result.
3. From this subtract p.
4. Take the square root of the result.
5. Add this to half of s; this is one of the two numbers and the other is s minus this one.

For example, if the sum is 10 and the product is 21 then the successive steps give 5, 25, 4, 2, 7 and $10 - 7 = 3$. Thus the two numbers are 7 and 3.

That this normal form is indeed a quadratic equation can be seen by multiplying the equation $s = x + y$ by x to find $sx = x^2 + xy = x^2 + p$. In other words, x is a solution of the quadratic equation $x^2 - sx + p = 0$ and, by symmetry, so is y.

Conversely, the solution of any quadratic equation can *in our notation* be viewed as the solution of a problem in normal form. Specifically, the equation $ax^2 + bx + c = 0$ can be rewritten as $x^2 + (c/a) = -(b/a)x$ and the solution of this equation is equivalent to finding two numbers whose sum is $-b/a$ and whose product is c/a. The Babylonians could *not* reduce *all* quadratic equations to a single normal form, however, because their arithmetic did not include negative numbers. To deal with this fact, they had a second normal form, in which the *difference* and the product of two numbers were given. This is a technical problem of considerable historical interest—it was only a few centuries ago that negative numbers became generally accepted so that polynomial equations did not have to be divided into several cases depending on the signs of the coefficients—but is of no importance to the algebra of the problem and will not be considered further here.

* The texts cannot be closely dated. Neugebauer places them between 1600 and 1800 B.C.

In modern algebraic notation (also only a few centuries old) the Babylonian solution of the problem in normal form can be written

$$x = \sqrt{\left(\frac{s}{2}\right)^2 - p} + \frac{s}{2}, \qquad y = s - x,$$

or, in a more familiar form,

$$x, y = \frac{s \pm \sqrt{s^2 - 4p}}{2}.$$

Thus it is fair to say that they knew the quadratic formula but that they spelled out the steps of the procedure instead of expressing it as a formula in the way we do.

How did they derive this procedure? Unfortunately, there is no indication in the texts which survive. The point of these texts seems to have been to convey, by means of several worked examples, the technique of solution. It is entirely possible that the technique was discovered by an ancient genius and that his successors merely adopted it because it produced correct answers. On the other hand, it may be that some derivation was well understood by many people at the time, but was transmitted orally or does not happen to be among the texts that have been found.

Cubic and Quartic Equations A.D. 1500

§4 There was some progress in algebra in the 3000 years between the Old Babylonian period and the Italian Renaissance, but it was not great. The late Greek writer Diophantus (circa A.D. 250) introduced some abbreviated algebraic notation, the Hindus used negative numbers on occasion, and the Arabs constructed the solutions of cubic equations as points of intersection of conic sections. When the Renaissance came, however, the advances in algebra were enormous, and they opened the way to great progress in all branches of mathematics.

In mathematics, the Renaissance was not a rebirth at all, but a period of first vigorous growth. In ancient times, Europe had been a mathematical backwater, and even the Romans were barbarians when it came to mathematics. During the Middle Ages, Europe had learned about algebra (al-jabr) from the Arabs and had begun to improve it by devising new symbols and notations. Then, in the sixteenth century, an enormous advance was made—the algebraic solution of cubic equations was discovered, and soon thereafter the solution of quartic equations.

The history of the discovery of these solutions and their exact description in terms of the still quite clumsy notation of the period will not be necessary in what follows. Instead, we will give just a brief account, in modern notation, of the solutions themselves. (For more details see Kline [K2], pp. 263–270 and 282–284.)

§5 Suppose the cubic equation to be solved has the form $x^3 + px + q = 0$. (An arbitrary cubic equation can be put in this form by dividing by the

coefficient of x^3 and then taking a change of variable $x' = x - c$ with c equal to the coefficient of x^2 divided by 3.) Introduce two new variables a and b and set $x = a - b$. The desired equation is then $a^3 - 3a^2b + 3ab^2 - b^3 + pa - pb + q = 0$, that is, $a^3 - b^3 + (a - b)(-3ab + p) + q = 0$. If it is stipulated that $3ab = p$, then this equation takes the form $a^3 - b^3 + q = 0$. If a solution (a, b) of these two equations $3ab = p$ and $a^3 - b^3 + q = 0$ in two unknowns* can be found, then, as is easily checked, the quantity $x = a - b$ is a solution of the original equation $x^3 + px + q = 0$. Multiplication by 3^3a^3 makes it possible to eliminate b from $a^3 - b^3 + q = 0$ to find $27a^6 - (3ab)^3 + 27a^3q = 0$, that is, $27a^6 + 27qa^3 - p^3 = 0$. This is a quadratic equation for a^3. Let a be the cube root of a solution of this quadratic equation and let $b = p/3a$. Then $3ab = p$ and $a^3 - b^3 + q = 0$, which implies that $x = a - b$ is a solution of the given equation.

§6 For the solution of the quartic, assume that the equation has the form $x^4 + px^2 + qx + r = 0$. (Again, an arbitrary quartic equation can easily be put in this form.) Let this be put in the form $x^4 = -px^2 - qx - r$. Then, if a is a new variable, $(x^2 + a)^2 = x^4 + 2ax^2 + a^2 = (-p + 2a)x^2 - qx + (-r + a^2)$. In order to take a square root on the right side, this quadratic function of x should have a single root—i.e. should be of the form $A(x + B)^2$—and by the quadratic formula this occurs if and only if $q^2 - 4(-p + 2a)(-r + a^2) = 0$. This is a cubic equation for a, which can (by the above method) be solved for a. When a is a root of this equation, the right side of the above expression of $(x^2 + a)^2$ has the form $A(x + B)^2$ where A is the coefficient of x^2 and B is the coefficient of x divided by $2A$, that is,

$$(x^2 + a)^2 = (-p + 2a)\left(x - \frac{q}{2(-p + 2a)}\right)^2,$$

or, more simply,

$$x^2 + a = \pm\sqrt{-p + 2a}\left(x - \frac{q}{2(-p + 2a)}\right),$$

which gives x as the solution of a quadratic equation.

§7 Of course the successful solution of the cubic and quartic equations led to attempts to solve the quintic equation. It was not until almost 300 years later, in the 1820's, that it was shown, first by Abel, then by Galois, that it is *impossible* to solve the quintic equation in the same manner that the cubic and the quartic were solved, specifically, by using no operations other than addition, subtraction, multiplication, division, and the extraction of roots.

During these 300 years the fruitful developments in algebra were in other

* There is no sharp distinction made here among the terms "variable", "unknown", and "indeterminate". For the most part, "variable" is used in this book. If a variable occurs in an equation that is to be solved, it may be called an unknown. If it is to remain variable and is being used primarily as a placeholder in a computation, it may be called an indeterminate.

directions. One of the most important was a theorem discovered by Isaac Newton, which is the subject of the next section.

Newton and Symmetric Functions

§8 Isaac Newton (1643–1727) is most famous for his discovery of the universal law of gravitation and for his use of that law to give an exact mathematical description of planetary motion. Consequently, he is identified in most people's minds with mathematical physics and applied mathematics. Even people who have some acquaintance with the history of mathematics and who realize that Newton, with Leibniz, is regarded as the father of differential and integral calculus, tend to think of Newton's mathematics as being closely related to his physics, and his calculus as being primarily a tool in his study of motion. Nevertheless, Newton's contributions to pure mathematics alone are sufficient to place him among the greatest geniuses in the history of mathematics. This section is devoted to a theorem of pure algebra which is of crucial importance to the later development of the subject and which appears to be Newton's creation.

A portion of this theorem was published in Newton's *Arithmetica Universalis* in 1707, after Newton was world famous and had ceased active scientific work. It is cited by Gauss [G2, Art. 338] and Weber [W3, vol. I, Sec. 46], among others, and is generally known as Newton's theorem. Of course the *Arithmetica Universalis* was known to have been written long before 1707, but it is only with the recent work of Derek T. Whiteside in analyzing, annotating, and publishing Newton's notebooks and papers that it has been possible to date many of Newton's discoveries and, in the case of the theorem under discussion, to know that he was aware at a very early date of the *full* theorem, not just the portion given in the *Arithmetica Universalis*.

§9 Whiteside found in papers dating to 1665–1666, in the very earliest phase of Newton's career, the following formulas: Let r, s, t be the three roots of a cubic equation $x^3 + bx^2 + cx + d = 0$, and let an expression like "every $r^i s^j$" denote the sum of all distinct expressions of the form $r^i s^j$ where r and s are roots of the given cubic, i.e. "every $r^2 s$" $= r^2 s + s^2 t + t^2 r + r^2 t + t^2 s + s^2 r$, "every r^2" $= r^2 + s^2 + t^2$, "every $r^2 s^2 t^2$" $= r^2 s^2 t^2$, etc. Then Newton's formulas* are

$$(\text{every } r) = -b \tag{1}$$

$$(\text{every } r^2) = b^2 - 2c \tag{2}$$

$$(\text{every } r^3) = -b^3 + 3bc - 3d \tag{3}$$

$$(\text{every } rs) = c \tag{4}$$

$$(\text{every } r^2 s) = -bc + 3d \tag{5}$$

* [N3, p. 517]. Newton took $-r, -s, -t$ to be the roots of the equation, which simply changes the signs of the formulas with odd degree.

$$(\text{every } r^3s) = b^2c - 2c^2 - bd \tag{6}$$
$$(\text{every } r^2s^2) = c^2 - 2bd \tag{7}$$
$$(\text{every } r^3s^2) = -bc^2 + 2b^2d + cd \tag{8}$$
$$(\text{every } r^3s^3) = c^3 - 3bcd + 3d^2 \tag{9}$$
$$(\text{every } rst) = -d \tag{10}$$
$$(\text{every } r^2st) = bd \tag{11}$$
$$(\text{every } r^3st) = -b^2d + 2cd \tag{12}$$
$$(\text{every } r^2s^2t) = -cd \tag{13}$$
$$(\text{every } r^3s^2t) = bcd - 3d^2 \tag{14}$$
$$(\text{every } r^3s^3t) = -c^2d + 2bd^2 \tag{15}$$
$$(\text{every } r^2s^2t^2) = d^2 \tag{16}$$
$$(\text{every } r^3s^2t^2) = -bd^2 \tag{17}$$
$$(\text{every } r^3s^3t^2) = cd^2 \tag{18}$$
$$(\text{every } r^3s^3t^3) = -d^3 \tag{19}$$

He did not record in his notes the method by which he derived these formulas, and we can only guess what lay behind them. However, it seems likely that the choice to stop with third powers of the roots was arbitrary* and that he could have given analogous formulas for higher powers. Moreover, the decision to deal with the three roots of a cubic, rather than the four roots of a quartic or the five roots of a quintic, was also probably arbitrary. In fact, a few pages later in Whiteside's book, a passage from Newton's notebook is reproduced in which he gives the analogs of formulas (1)–(3) for an equation of degree 8, namely, the formulas†

$$(\text{every } r) = -p,$$
$$(\text{every } r^2) = p^2 - 2q,$$
$$(\text{every } r^3) = -p^3 + 3pq - 3r,$$
$$(\text{every } r^4) = p^4 - 4p^2q + 4pr - 4s + 2q^2,$$
$$(\text{every } r^5) = -p^5 + 5p^3q - 5p^2r + 5ps - 5t - 5pq^2 + 5qr,$$
$$(\text{every } r^6) = p^6 - 6p^4q + 6p^3r - 6p^2s + 6pt - 6v + 9p^2q^2 - 12pqr + 6qs - 2q^3,$$
$$(\text{every } r^7) = -p^7 + 7p^5q - 7p^4r + 7p^3s - 7p^2t + 7pv - 7y,$$
$$(\text{every } r^8) = p^8 - 8p^6q + 8p^5r - 8p^4s + 8p^3t - 8p^2v + 8py - 8z,$$

* Newton in fact had a specific goal in mind in the passage in question, namely, the derivation of the explicit formula for the resultant of two cubics (see Exercise 8). For this goal he needed the given formulas and only these.

† The first four of these formulas were published by Albert Girard in 1692. In Whiteside's opinion, Newton was not aware of Girard's work.

where r runs over the eight roots of the 8th degree equation $x^8 + px^7 + qx^6 + rx^5 + sx^4 + tx^3 + vx^2 + yx + z = 0$.

In other words, it appears likely that Newton was aware that there are analogous formulas for all degrees, that is, that *any symmetric polynomial in the roots of an equation can be expressed in terms of the coefficients of that equation.* This theorem is the foundation stone of Galois theory, so it is important to have a careful statement and proof of it before proceeding. (It must be admitted, however, that neither a careful statement nor a proof of it seems to have been published before the nineteenth century. Everyone seemed familiar with it and used it without inhibition.)

The Fundamental Theorem on Symmetric Polynomials

§10 The first step in giving a careful statement of the theorem is to remove the reference to roots of an nth degree equation, because these roots may be irrational or complex and they are really extraneous to the theorem. (Newton explicitly states in his formulas that the roots may be "false", i.e. negative, or "imaginary".) The particular formulas (1), (4) and (10) in Newton's list, that is,

$$\begin{aligned} r + s + t &= -b, \\ rs + st + tr &= c, \\ rst &= -d, \end{aligned} \tag{20}$$

are especially important and were probably rather widely known in Newton's time. (Whiteside [N3, p. 518, note 12] observes that the general case of these formulas was published by Albert Girard in 1629 but says that "we may assume" that Newton's version of it, which he published in the *Arithmetica Universalis*, was his "independent discovery".) These formulas follow immediately from the identity

$$x^3 + bx^2 + cx + d = (x - r)(x - s)(x - t),$$

when the right side is multiplied out and coefficients of like powers of x are equated. The same procedure applied to

$$x^n + b_1x^{n-1} + b_2x^{n-2} + \cdots + b_n = (x - r_1)(x - r_2)\cdots(x - r_n)$$

shows that, in analogy to (20), the sum of all* $\binom{n}{k}$ products of k of the r_i is equal to $(-1)^k b_k$. That is,

$$\begin{aligned} r_1 + r_2 + \cdots + r_n &= -b_1, \\ r_1r_2 + r_1r_3 + \cdots + r_{n-1}r_n &= b_2, \\ r_1r_2r_3 + r_1r_2r_4 + \cdots + r_{n-2}r_{n-1}r_n &= -b_3, \\ &\vdots \\ r_1r_2\cdots r_n &= (-1)^n b_n. \end{aligned}$$

* Here $\binom{n}{k}$ denotes the binomial coefficient $\dfrac{n!}{k!(n-k)!}$.