2001

Modular Miracles

John Stillwell
University of San Francisco, stillwell@usfca.edu

Recommended Citation

John Stillwell. Modular Miracles. The American Mathematical Monthly. Vol. 108, No. 1 (Jan., 2001), pp. 70-76. DOI: 10.2307/2695682

THE EVOLUTION OF ...

Edited by **Abe Shenitzer and John Stillwell**

Modular Miracles

John Stillwell

Over the last 20 years, the modular function has become widely known through its miraculous intervention in two great mathematical achievements: the proof of Fermat's last theorem and the "moonshine" of the monster simple group. In both cases, the modular function appears where no one expected it, and it bridges a chasm between seemingly unrelated fields. It is probably fair to say that, in these two cases, we do not yet fully understand how the modular magic works.

However, it can at least be said that these are not the first modular miracles. Ever since its discovery, in the early 19th century, the modular function has been an engine for spectacular and unexpected results. Now that things modular are back in the news, it is a good time to recall some of the modular miracles of the 19th century. They help us see the recent results in some perspective, and encourage us to believe that there is a lot more to be learned.

THE MODULAR FUNCTION *j*. Modular functions may be defined as meromorphic functions on the upper half plane with the periodicity of the *modular tessellation* shown in Figure 1. When the half plane is interpreted as the hyperbolic plane, the black and white tiles of the tessellation are congruent triangles with one vertex at infinity, and the whole tessellation is generated by reflections in the sides of any one of them.

It follows that a modular function is determined by its values on any one tile of the tessellation, the other values being obtained by reflection in the sides of the tile. The values on a tile can be defined by mapping the tile conformally onto the upper half plane, and they in turn are completely determined by the images of the three vertices.

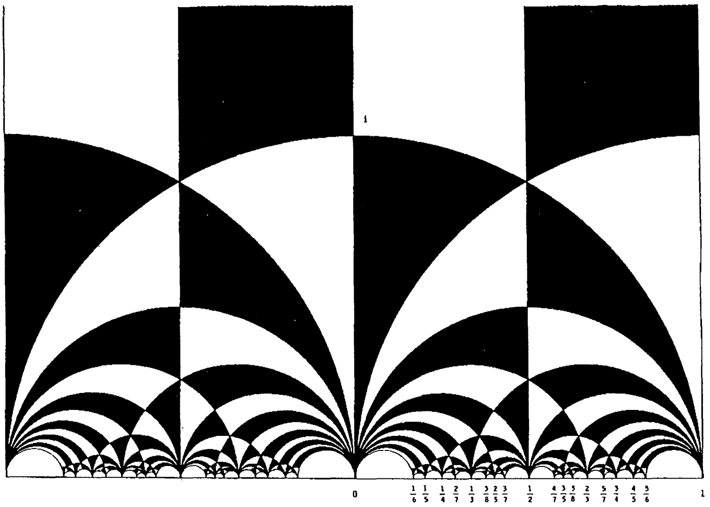

Figure 1. The modular tessellation

This idea was used by Dedekind (1877) to define the classical modular function j by the unique conformal map

$$\text{white region} \rightarrow \text{half plane}$$

which sends $i, e^{\pi i/3}, \infty$ to $0, 1, \infty$ respectively [3].

The periodicity of j can be described algebraically by saying that

$$j(\tau) = j\left(\frac{a\tau + b}{c\tau + d}\right)$$

for any $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$. The transformations

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}$$

carry any particular black and white region to any other. These transformations are generated by the two simple transformations $\tau \mapsto \tau + 1$ and $\tau \mapsto -1/\tau$, so the latter transformations also define the periodicity of j.

Because of its periodicity under $\tau \mapsto \tau + 1$, j has a Fourier series, that is, an expansion in powers of $q = e^{2i\pi\tau}$. This expansion happens to be

$$j(\tau) = q^{-1} + 744 + 196884q + 21493760q^2 + \cdots$$

Neither the tessellation definition nor the expansion in powers of q is close to the original definition of j, which comes from the theory of elliptic functions, as we explain below. We began with the geometric definition because it is probably the simplest to grasp, though it hides some difficulties (among them the Riemann mapping theorem, which ensures the *existence* of a conformal map of any simply-connected region onto the half plane). The definition of j as a mapping also yields the most famous application of the modular function to analysis—Picard's proof that any entire function omits at most one complex value. We do not go further into Picard's theorem, because it can be found in most complex analysis books; Picard's beautiful proof is presented in [1, p. 307].

For a thorough treatment of j and its history, including most of the topics discussed in this article, McKean and Moll's book [8] is warmly recommended.

THE QUINTIC MIRACLE. *The general quintic equation can be solved by* j. This result was proved by Hermite in 1858 [5]. It was not completely out of the blue, because Galois had pointed out a quintic equation related to j in 1832, and Kronecker had similar ideas about the same time as Hermite. Nevertheless, it is a startling result, and it remains so even when its antecedents are pointed out.

Hermite compared his solution of the quintic by j to the solution of the cubic equation that takes advantage of the "angle-tripling" equation

$$4\cos^3\theta - 3\cos\theta = \cos 3\theta$$

satisfied by the cosine function. One transforms the general cubic equation into the special form

$$4x^3 - 3x = c,$$

and then sets $x = \cos\theta$, where $c = \cos 3\theta$.

There are analogous *modular equations* satisfied by j, and it turns out that the general quintic equation can be transformed to the quintic modular equation.

Where Do Modular Equations Come From? The function j is not the only function with the periodicity of the modular tessellation, but it is simplest in the sense that all other such functions are rational functions of j. The first of them to be encountered, and the origin of the name "modular", was *modulus* k^2 in the elliptic integral

$$\int \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}.$$

A thought-provoking result about such integrals was Fagnano's 1718 formula for doubling the arc length of the lemniscate:

$$2\int_0^x \frac{dt}{\sqrt{1-t^4}} = \int_0^y \frac{dt}{\sqrt{1-t^4}}, \text{ where } y = \frac{2x\sqrt{1-x^4}}{1+x^4}$$

which gives a polynomial equation between x and y:

$$y^2(1+x^4)^2 = 4x^2(1-x^4).$$

This is analogous to the doubling formula for the arcsine integral,

$$2\int_0^x \frac{dt}{\sqrt{1-t^2}} = \int_0^y \frac{dt}{\sqrt{1-t^2}}, \text{ where } y = 2x\sqrt{1-x^2},$$

which in turn is just a restatement of the double angle formula

$$\sin 2\theta = 2\sin\theta\cos\theta = 2\sin\theta\sqrt{1-\sin^2\theta},$$

and the polynomial relation

$$y^2 = 4x^2(1-x^2)$$

between $y = \sin 2\theta$ and $x = \sin\theta$. This analogy with circular functions led to great interest in n-tupling (and later, multiplication by complex numbers, or "complex multiplication") of elliptic integrals and to computation of the corresponding polynomial equations. When the value of the integral is regarded as a function of the modulus, the equations obtained are called *modular equations.*

Modular equations were a popular topic with many leading mathematicians of the early 19th century—Legendre, Gauss, Abel, Jacobi, Galois—and the results of Galois were particularly tantalising. Galois left only some cryptic remarks about the equations for multiplication by 5, 7, and 11 (implying that they yield equations of degrees 5, 7, and 11) in the letter he wrote to Chevalier just before his death. It was several decades before these remarks were really understood, and Hermite's 1858 paper was both a step towards understanding Galois, and a step beyond him.

THE QUADRATIC MIRACLE. Kronecker (1857) discovered that *j detects the class number of $\mathbb{Q}(\sqrt{-D})$ for an imaginary quadratic integer $\sqrt{-D}$* [7]. This result is to my mind even more startling than the solution of the quintic, because class numbers are a deeper topic, which mathematicians did not begin to grasp until the 1830s.

In 1832 Gauss studied the *Gaussian integers $a + ib$*, where $a, b \in \mathbb{Z}$ and $i = \sqrt{-1}$, and showed that they have unique prime factorisation or *class number 1*. (The terminology goes back to the older language of quadratic forms, where the equivalent fact in this case is that all forms $ax^2 + bxy + cy^2$ with $b^2 - 4ac = -4$ are in the same "class" as $x^2 + y^2$.) Soon afterwards, mathematicians noticed examples, such as the quadratic integers $a + b\sqrt{-5}$, where prime factorisation is *not* unique because the class number is > 1. (In this case the class number is 2, and the two classes of forms are represented by $x^2 + 5y^2$ and $2x^2 + 2xy + 3y^2$.) In 1839, Dirichlet introduced the powerful analytic method of Dirichlet series to determine the class number of the integers of a quadratic field $\mathbb{Q}(\sqrt{-D})$, but it was a complete surprise when Kronecker showed that j could do the same job.

He showed that, for any integer τ in the quadratic field $\mathbb{Q}(\sqrt{-D})$, *$j(\tau)$ is an algebraic integer whose degree is the class number of $\mathbb{Q}(\sqrt{-D})$.*

For example, the Gaussian integers are the integers of the field $\mathbb{Q}(i)$ with $D = 1$, and it turns out that $j(i) = 12^3$—an ordinary integer, as we expect because $\mathbb{Q}(i)$ has class number 1. A second example, which happens to be the largest D for which $\mathbb{Q}(\sqrt{-D})$ has class number 1, is where $D = 163$. Gauss also found this example, and the class number 1 is confirmed by the ordinary integer value

$$j((1 + \sqrt{-163})/2) = (-640320)^3.$$

Finally, an example with class number 2 is $\mathbb{Q}(\sqrt{-15})$, and indeed

$$j((1 + \sqrt{-15})/2) = (-191025 + 85995\sqrt{5})/2,$$

which is an integer of degree 2. (The existence of "integers" with denominator 2 is a quirk of certain quadratic fields that the reader may take on trust here.)

Kronecker's result is difficult to explain in a short article, but we can give the following hint. What the quadratic integers have in common with elliptic functions is an underlying *lattice* L in the plane $\mathbb{C}$—a set of points at the corners of a tessellation of the plane by identical parallelograms. An elliptic function f has two *periods* ω_1 and ω_2, which have different directions in $\mathbb{C}$ and hence generate the *lattice of periods* $\{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}$, at each point of which f takes the same value. In a quadratic field $\mathbb{Q}(\sqrt{-D})$ the set $\mathcal{O}$ of *integers* is a lattice, either

$$\{m + n\sqrt{-D} : m, n \in \mathbb{Z}\} \quad \text{or} \quad \left\{\frac{m}{2} + \frac{n\sqrt{-D}}{2} : m, n \in \mathbb{Z} \text{ with same parity}\right\},$$

and more generally so is any *ideal* of $\mathcal{O}$—a set of integers closed under addition and under multiplication by any $\alpha \in \mathcal{O}$. The algebraic significance of ideals is that $\mathcal{O}$ has unique prime factorisation if and only if every ideal of $\mathcal{O}$ is *principal*—that is, equal to $\alpha\mathcal{O}$ for some $\alpha \in \mathcal{O}$—and a principal ideal $\alpha\mathcal{O}$ is geometrically significant because it has the *same shape* as $\mathcal{O}$ (being the result of magnifying $\mathcal{O}$ by $|\alpha|$ and rotating by $\arg \alpha$).

The modular function is pertinent to both elliptic functions and quadratic integers because *j is really a function of lattice shapes*. The idea of lattice shape may be illustrated by the lattice of periods $\{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}$. The points of this lattice occur at the corners of the tessellation of the plane by parallelograms shown in Figure 2.

The "shape" of a parallelogram is captured by the number $\omega = \omega_2/\omega_1$, because $|\omega|$ is the ratio $|\omega_2|/|\omega_1|$ of the side lengths and $\arg \omega = \arg \omega_2 - \arg \omega_1$ is the angle between the sides. However, this parallelogram is just one of infinitely many that define the same lattice. Another is shown in Figure 3.

The shape of the basic parallelogram is now

$$\frac{\omega_2 + \omega_1}{\omega_1} = \omega + 1,$$

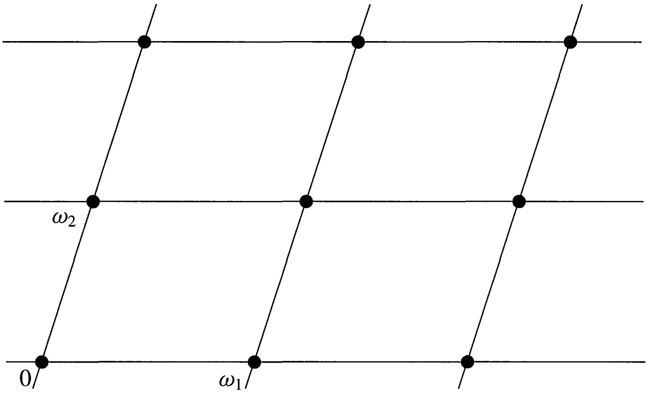

Figure 2. Parallelograms generated by ω_1 and ω_2

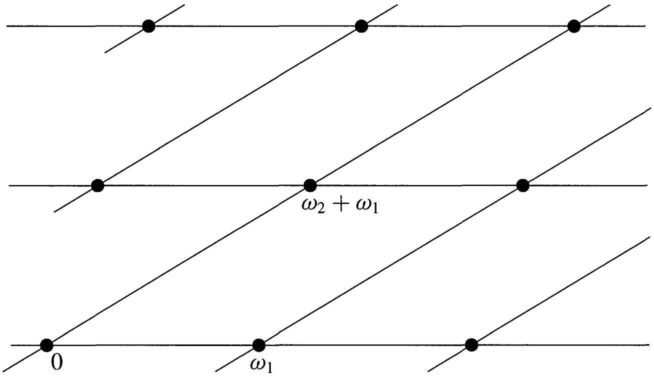

Figure 3. Parallelograms generated by ω_1 and $\omega_2 + \omega_1$

so the lattice shape is represented equally well by $\omega + 1$. It turns out that, for each ω representing the shape of a lattice L, the number $(a\omega + b)/(c\omega + d)$ also represents the shape of L, provided $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$. A *lattice shape* is therefore a whole *class* of numbers, of the form

$$\frac{a\omega + b}{c\omega + d} \quad \text{for some } \omega,$$

where $a, b, c, d \in \mathbb{Z}$ and $ad - bc = 1$.

This brings us to the reason for saying that j is a function of lattice shapes: as mentioned at the beginning, j has the property that

$$j(\omega) = j\left(\frac{a\omega + b}{c\omega + d}\right)$$

for any $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$. Thus j *takes the same value at each number in a lattice shape.* In other words, j is *well defined* on lattice shapes.

Now the properties of an elliptic function are largely controlled by the shape of its period lattice, and the properties of a quadratic field $\mathbb{Q}(\sqrt{-D})$ are controlled by the shapes of its ideals. In particular, it turns out that $\mathbb{Q}(\sqrt{-D})$ *has unique prime factorisation if and only if all its ideals have the same shape.* This is why j can have something to say about unique prime factorisation in quadratic fields—it is an echo of what j says about the so-called "complex multiplication" of period lattices—though the way j says it is still pretty amazing.

The equation satisfied by $j(\sqrt{-D})$ happens to be another modular equation, which factorises into terms like $x - j(\sqrt{-D})$, and the number of factors is the number of different lattice shapes in the integers of $\mathbb{Q}(\sqrt{-D})$

THE NUMERICAL MIRACLE. Hermite (1859) noticed a curious numerical consequence of Kronecker's theorem on the values of $j(\tau)$:

$$e^{\pi\sqrt{163}} = 262537412640768744,$$

(an integer!) correct to 12 decimal places [6].

This little known discovery of Hermite was exploited by Martin Gardner in an amusing hoax edition of his column in *Scientific American.* On 1 April 1975 Gardner announced—among several other "sensational discoveries that have somehow or another escaped public attention"—that

$$e^{\pi\sqrt{163}} = 262537412640768744 \text{ exactly.} \tag{1}$$

He gave the announcement an extra coat of varnish by claiming that it settled a conjecture of Ramanujan, supposedly made in a paper of 1914. The paper cited by Gardner does indeed discuss near integers of the form $e^{\pi\sqrt{n}}$, but without claiming that they could be integers, and without mentioning $e^{\pi\sqrt{163}}$. Still, in the pocket calculator days of 1975, it was pretty hard to decide whether $e^{\pi\sqrt{163}}$ is an integer or not.

Its true value, as Hermite and Ramanujan knew, is the integer in (1) *minus a very tiny number* $(< 10^{-12})$.

In fact, putting $\tau = (1 + \sqrt{-163})/2$ in $q = e^{2i\pi\tau}$ gives the tiny

$$q = e^{i\pi - \pi\sqrt{163}} = -e^{-\pi\sqrt{163}},$$

and putting this q in

$$j(\tau) = q^{-1} + 744 + 196884q + 21493760q^2 + \cdots$$

gives

$$j((1 + \sqrt{-163})/2) = -e^{\pi\sqrt{163}} + 744 - \text{tiny number},$$

and therefore

$$e^{\pi\sqrt{163}} = \text{integer} - \text{tiny number}.$$

THE ORIGIN OF MOONSHINE. "Moonshine" is a theory linking j to the monster simple group $\mathbb{M}$, and it has its origin in an apparent coincidence observed by McKay in 1977: *The coefficient* 196884 *in the Fourier expansion of* j *is* 1 *plus the dimension of the smallest nontrivial representation of* $\mathbb{M}$.

Actually, several other coincidences were discovered around the same time, and are listed in [2]. But nonetheless moonshine could hardly have been discovered without knowing the Fourier expansion of j, so one would like to know who discovered the coefficient 196884. Hermite 1859 actually has the *incorrect* expansion

$$j(\tau) = q^{-1} + 744 + 196880q + \cdots,$$

though the error does not affect his result that $e^{\pi\sqrt{163}}$ is an integer to 12 decimal places.

As far as I know, the first correct expansion as far as the coefficient 196884 was given by Weber in 1891 [9, p. 248]. Was this the first glimpse of moonshine? Or did Hermite also see 196884, but write it down incorrectly? I am inclined to vote for Hermite because his 1859 paper contains another series that later became part of moonshine [2, p. 334]:

$$q^{-1} + 104 + 4372q + 96256q^2 + \cdots$$

and in this series Hermite got all the digits right.

REFERENCES

1. L. V. Ahlfors, *Complex Analysis*, McGraw-Hill Kogakusha, Tokyo, 1979.
2. J. H. Conway and S. P. Norton, Monstrous moonshine, *Bull. London Math. Soc.* 11 (1979) 308–339.
3. R. Dedekind, Schreiben an Herrn Borchardt über die Theorie der elliptischen Modulfunktionen, *J. Reine Angew. Math.* 83 (1877) 265–292.
4. M. Gardner, Mathematical Games, *Scientific American* 232 (April 1975), p. 126.
5. C. Hermite, Sur la résolution de l'équation du cinquième degré, *C. R. Acad. Sci. Paris Sér. I Math.* XLVI (1858) 508. Also in *Oeuvres de Charles Hermite*, Paris, Gauthier-Villars, 1905-17, vol. 2, pp. 5–12. .
6. C. Hermite, Sur la théorie des équations modulaires, *C. R. Acad. Sci. Paris Sér. I Math.* XLVIII (1859) 940. Also in *Oeuvres de Charles Hermite*, Paris, Gauthier-Villars, 1905-17, vol. 2, pp. 38–82.
7. L. Kronecker, Über die elliptischen Functionen für welche complexe Multiplication stattfindet, *Leopold Kronecker's Werke*, Chelsea Pub. Co., New York, 1968, vol. 4, pp. 179–183.
8. H. P. McKean and V. Moll, *Elliptic Curves*, Cambridge University Press, Cambridge, 1997.
9. Weber, *Elliptische Functionen und algebraische Zahlen*, Vieweg, Braunschweig, 1891.

Monash University, Clayton, Australia
john.stillwell@monash.edu.au