0

Introduction: glimpses of the theory beneath Monstrous Moonshine

> When you are collecting mushrooms, you only see the mushroom itself. But if you are a mycologist, you know that the real mushroom is in the earth. There's an enormous thing down there, and you just see the fruit, the body that you eat. In mathematics, the upper part of the mushroom corresponds to theorems that you see, but you don't see the things that are below, that is: *problems*, *conjectures*, *mistakes*, *ideas*, etc.
>
> *V. I. Arnold [**17**]*

> What my experience of mathematical work has taught me again and again, is that the proof always springs from the insight, and not the other way around – and that the insight itself has its source, first and foremost, in a delicate and obstinate feeling of the relevant entities and concepts and their mutual relations. The guiding thread is the inner coherence of the image which gradually emerges from the mist, as well as its consonance with what is known or foreshadowed from other sources – and it guides all the more surely as the 'exigence' of coherence is stronger and more delicate.
>
> *A. Grothendieck.*[1]

Interesting events (e.g. wars) always happen whenever different realisations of the same thing confront one another. When clarity and precision are added to the mix, we call this mathematics. In particular, the most exciting and significant moments in mathematics occur when we discover that seemingly unrelated phenomena are shadows cast by the same beast. This book studies one who has been recently awakened.

In 1978, John McKay made an intriguing observation: $196\,884 \approx 196\,883$. *Monstrous Moonshine* is the collection of questions (and a few answers) that it directly inspired. No one back then could have guessed the riches to which it would lead. But in actual fact, Moonshine (albeit non-Monstrous) really began long ago.

0.1 Modular functions

Up to topological equivalence (homeomorphism), every compact surface is uniquely specified by its genus: a sphere is genus 0, a torus genus 1, etc. However, a (real) surface can be made into a complex curve by giving it more structure. For a sphere, up to

[1] Translated in *Geometric Galois Actions* 1, edited by L. Schneps *et al.* (Cambridge, Cambridge University Press, 1997) page 285.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

complex-analytic equivalence there is only one way to do this, namely the Riemann sphere $\mathbb{C} \cup \{\infty\}$. Surfaces of genus > 0 can be given complex structure in a continuum of different ways.

Any such complex curve Σ is complex-analytically equivalent to one of the form $\Gamma\backslash\overline{\mathbb{H}}$. The *upper half-plane*

$$\mathbb{H} := \{\tau \in \mathbb{C} \mid \operatorname{Im} \tau > 0\} \tag{0.1.1}$$

is a model for hyperbolic geometry. Its geometry-preserving maps form the group $\mathrm{SL}_2(\mathbb{R})$ of 2×2 real matrices with determinant 1, which act on $\mathbb{H}$ by the familiar

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} . \tau = \frac{a\tau + b}{c\tau + d}. \tag{0.1.2}$$

Γ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$. By $\overline{\mathbb{H}}$ here we mean $\mathbb{H}$ with countably many points from its boundary $\mathbb{R} \cup \{\mathrm{i}\infty\}$ added – these extra boundary points, which depend on Γ, are needed for $\Gamma\backslash\overline{\mathbb{H}}$ to be compact. The construction of the space $\Gamma\backslash\overline{\mathbb{H}}$ of Γ-orbits in $\overline{\mathbb{H}}$ is completely analogous to that of the circle $\mathbb{R}/\mathbb{Z}$ or torus $\mathbb{R}^2/\mathbb{Z}^2$. See Section 2.1.1 below.

The most important example is $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, because the moduli space of possible complex structures on a torus can be naturally identified with $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. For that Γ, as well as all other Γ we consider in this book, we have

$$\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{Q} \cup \{\mathrm{i}\infty\}. \tag{0.1.3}$$

These additional boundary points $\mathbb{Q} \cup \{\mathrm{i}\infty\}$ are called cusps.

Both geometry and physics teach us to study a geometric shape through the functions (fields) that live on it. The functions f living on $\Sigma = \Gamma\backslash\overline{\mathbb{H}}$ are simply functions $f : \overline{\mathbb{H}} \to \mathbb{C}$ that are periodic with respect to Γ: that is,

$$f(A.\tau) = f(\tau), \qquad \forall \tau \in \overline{\mathbb{H}},\ A \in \Gamma. \tag{0.1.4}$$

They should also preserve the complex-analytic structure of Σ. Ideally this would mean that f should be holomorphic but this is too restrictive, so instead we require meromorphicity (i.e. we permit isolated poles).

Definition 0.1 *A* modular function *f for some Γ is a meromorphic function $f : \overline{\mathbb{H}} \to \mathbb{C}$, obeying the symmetry* (0.1.4).

It is clear then why modular functions must be important: they are the functions living on complex curves. In fact, modular functions and their various generalisations hold a central position in both classical and modern number theory.

We can construct some modular functions for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ as follows. Define the *(classical) Eisenstein series* by

$$G_k(\tau) := \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} (m\tau + n)^{-k}. \tag{0.1.5}$$

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

For odd k it identically vanishes. For even $k > 2$ it converges absolutely, and so defines a function holomorphic throughout $\mathbb{H}$. It is easy to see from (0.1.5) that

$$G_k\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^k\, G_k(\tau), \qquad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \tag{0.1.6}$$

and all τ. This transformation law (0.1.6) means that G_k isn't quite a modular function (it's called a *modular form*). However, various homogeneous rational functions of these G_k *will* be modular functions for $\mathrm{SL}_2(\mathbb{Z})$ – for example, $G_8(\tau)/G_4(\tau)^2$ (which turns out to be constant) and $G_4(\tau)^3/G_6(\tau)^2$ (which doesn't). All modular functions of $\mathrm{SL}_2(\mathbb{Z})$ turn out to arise in this way.

Can we characterise all modular functions, for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ say? We know that any modular function is a meromorphic function on the compact surface $\Sigma = \mathrm{SL}_2(\mathbb{Z})\backslash\overline{\mathbb{H}}$. As we explain in Section 2.2.4, Σ is in fact a sphere. It may seem that we've worked very hard merely to recover the complex plane $\mathbb{C} \cong \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ and its familiar compactification the Riemann sphere $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\} \cong \mathrm{SL}_2(\mathbb{Z})\backslash\overline{\mathbb{H}}$, but that's *exactly* the point!

Although there are large numbers of meromorphic functions on the complex plane $\mathbb{C}$, the only ones that are also meromorphic at ∞ – the only functions meromorphic on the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ – are the rational functions $\frac{\text{polynomial in } z}{\text{polynomial in } z}$ (the others have essential singularities there). So if J is a change-of-coordinates (or *uniformising*) function identifying our surface Σ with the Riemann sphere, then J (lifted to a function on the covering space $\overline{\mathbb{H}}$) will be a modular function for $\mathrm{SL}_2(\mathbb{Z})$, and any modular function $f(\tau)$ will be a rational function in $J(\tau)$:

$$f(\tau) = \frac{\text{polynomial in } J(\tau)}{\text{polynomial in } J(\tau)}. \tag{0.1.7}$$

Conversely, any rational function (0.1.7) in J is modular. Thus J generates modular functions for $\mathrm{SL}_2(\mathbb{Z})$, in a way analogous to (but stronger and simpler than) how the exponential $e(x) = e^{2\pi \mathrm{i}\, x}$ generates the period-1 smooth functions f on $\mathbb{R}$: we can always expand such an f in the pointwise-convergent Fourier series $f(x) = \sum_{n=-\infty}^{\infty} a_n\, e(x)^n$.

There is a standard historical choice j for this uniformisation J, namely

$$\begin{aligned} j(\tau) &:= 1728\, \frac{20\, G_4(\tau)^3}{20\, G_4(\tau)^3 - 49\, G_6(\tau)^2} \\ &= q^{-1} + 744 + 196\,884\, q + 21\,493\,760\, q^2 + 864\,299\,970\, q^3 + \cdots \end{aligned} \tag{0.1.8}$$

where $q = \exp[2\pi \mathrm{i}\, \tau]$. In fact, this choice (0.1.8) is canonical, apart from the arbitrary constant 744. This function j is called the absolute invariant or *Hauptmodul* for $\mathrm{SL}_2(\mathbb{Z})$, or simply the *$j$-function*.

0.2 The McKay equations

In any case, one of the best-studied functions of classical number theory is the j-function. However, its most remarkable property was discovered only recently: McKay's

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

approximations $196\,884 \approx 196\,883$, $21\,493\,760 \approx 21\,296\,876$ and $864\,299\,970 \approx 842\,609\,326$. In fact,

$$196\,884 = 196\,883 + 1, \tag{0.2.1a}$$

$$21\,493\,760 = 21\,296\,876 + 196\,883 + 1, \tag{0.2.1b}$$

$$864\,299\,970 = 842\,609\,326 + 21\,296\,876 + 2 \cdot 196\,883 + 2 \cdot 1. \tag{0.2.1c}$$

The numbers on the left sides of (0.2.1) are the first few coefficients of the j-function. The numbers on the right are the dimensions of the smallest irreducible representations of Fischer–Griess's *Monster finite simple group* $\mathbb{M}$.

A *representation* of a group G is the assignment of a matrix $R(g)$ to each element g of G in such a way that the matrix product respects the group product, that is $R(g)\,R(h) = R(gh)$. The dimension of a representation is the size n of its $n \times n$ matrices $R(g)$.

The *finite simple groups* are to finite groups what the primes are to integers – they are their elementary building blocks (Section 1.1.2). They have been classified (see [**22**] for recent remarks on the status of this proof). The resulting list consists of 18 infinite families (e.g. the cyclic groups $\mathbb{Z}_p := \mathbb{Z}/p\mathbb{Z}$ of prime order), together with 26 exceptional groups. The Monster $\mathbb{M}$ is the largest and richest of these exceptionals, with order

$$\|\mathbb{M}\| = 2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71 \approx 8 \times 10^{53}. \tag{0.2.2}$$

Group theorists would like to believe that the classification of finite simple groups is one of the high points in the history of mathematics. But isn't it possible instead that their enormous effort has merely culminated in a list of interest only to a handful of experts? Years from now, could the Monster – the signature item of this list – become a lost bone in a dusty drawer of a forgotten museum, remarkable only for its colossal irrelevance?

With numbers so large, it seemed doubtful to McKay that the numerology (0.2.1) was merely coincidental. Nevertheless, it was difficult to imagine any deep conceptual relation between the Monster and the j-function: mathematically, they live in different worlds.

In November 1978 he mailed the 'McKay equation' (0.2.1a) to John Thompson. At first Thompson likened this exercise to reading tea leaves, but after checking the next few coefficients he changed his mind. He then added a vital piece to the puzzle.

0.3 Twisted #0: the Thompson trick

A nonnegative integer begs interpretation as the dimension of some vector space. Essentially, that was what McKay proposed. Let $\rho_0, \rho_1, \ldots$ be the irreducible representations of $\mathbb{M}$, ordered by dimension. Then the equations (0.2.1) are really hinting that there is an infinite-dimensional graded representation

$$V = V_{-1} \oplus V_1 \oplus V_2 \oplus V_3 \oplus \cdots \tag{0.3.1}$$

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

of $\mathbb{M}$, where $V_{-1} = \rho_0$, $V_1 = \rho_1 \oplus \rho_0$, $V_2 = \rho_2 \oplus \rho_1 \oplus \rho_0$, $V_3 = \rho_3 \oplus \rho_2 \oplus \rho_1 \oplus \rho_1 \oplus \rho_0 \oplus \rho_0$, etc., and the j-function is essentially its graded dimension:

$$j(\tau) - 744 = \dim(V_{-1})\, q^{-1} + \sum_{i=1}^{\infty} \dim(V_i)\, q^i. \tag{0.3.2}$$

Thompson [**525**] suggested that we twist this, that is more generally we consider what we now call the *McKay–Thompson series*

$$T_g(\tau) = \mathrm{ch}_{V_{-1}}(g)\, q^{-1} + \sum_{i=1}^{\infty} \mathrm{ch}_{V_i}(g)\, q^i \tag{0.3.3}$$

for each element $g \in \mathbb{M}$. The character 'ch_ρ' of a representation ρ is given by 'trace': $\mathrm{ch}_\rho(g) = \mathrm{tr}(\rho(g))$. Up to equivalence (i.e. choice of basis), a representation ρ can be recovered from its character ch_ρ. The character, however, is much simpler. For example, the smallest nontrivial representation of the Monster $\mathbb{M}$ is given by almost 10^{54} complex matrices, each of size $196\,883 \times 196\,883$, while the corresponding character is completely specified by 194 integers (194 being the number of 'conjugacy classes' in $\mathbb{M}$).

For any representation ρ, the character value $\mathrm{ch}_\rho(id.)$ equals the dimension of ρ, and so $T_{id.}(\tau) = j(\tau) - 744$ and we recover (0.2.1) as special cases. But there are many other possible choices of $g \in \mathbb{M}$, although conjugate elements g, hgh^{-1} have identical character values and hence have identical McKay–Thompson series $T_g = T_{hgh^{-1}}$. In fact, there are precisely 171 *distinct* functions T_g. Thompson didn't guess what these functions T_g would be, but he suggested that they too might be interesting.

0.4 Monstrous Moonshine

John Conway and Simon Norton [**111**] did precisely what Thompson asked. Conway called it 'one of the most exciting moments in my life' [**107**] when he opened Jacobi's foundational (but 150-year-old!) book on elliptic and modular functions and found that the first few terms of each McKay–Thompson series T_g coincided with the first few terms of certain special functions, namely the Hauptmoduls of various genus-0 groups Γ. Monstrous Moonshine – which conjectured that the McKay–Thompson series *were* those Hauptmoduls – was officially born.

We should explain those terms. When the surface $\Gamma\backslash\overline{\mathbb{H}}$ is a sphere, we call the group Γ *genus 0*, and the (appropriately normalised) change-of-coordinates function from $\Gamma\backslash\overline{\mathbb{H}}$ to the Riemann sphere $\mathbb{C} \cup \{\infty\}$ the *Hauptmodul* for Γ. All modular functions for a genus-0 group Γ are rational functions of this Hauptmodul. (On the other hand, when Γ has positive genus, two generators are needed, and there's no canonical choice for them.)

The word 'moonshine' here is English slang for 'insubstantial or unreal', 'idle talk or speculation',[2] 'an illusive shadow'.[3] It was chosen by Conway to convey as well the

[2] Ernest Rutherford (1937): 'The energy produced by the breaking down of the atom is a very poor kind of thing. Anyone who expects a source of power from the transformation of these atoms is talking moonshine.' (quoted in *The Wordsworth Book of Humorous Quotations*, Wordsworth Editions, 1998).

[3] *Dictionary of Archaic Words*, J. O. Halliwell, London, Bracken Books, 1987. It also defines moonshine as 'a dish composed partly of eggs', but that probably has less to do with Conway's choice of word.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

impression that things here are dimly lit, and that Conway and Norton were 'distilling information illegally' from the Monster character table.

In hindsight, the first incarnation of Monstrous Moonshine goes back to Andrew Ogg in 1975. He was in France discussing his result that the primes p for which the group $\Gamma_0(p)+$ has genus 0, are

$$p \in \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 41, 47, 59, 71\}.$$

(The group $\Gamma_0(p)+$ is defined in (7.1.5).) He also attended there a lecture by Jacques Tits, who was describing a newly conjectured simple group. When Tits wrote down the order (0.2.2) of that group, Ogg noticed its prime factors coincided with his list of primes. Presumably as a joke, he offered a bottle of Jack Daniels whisky to the first person to explain the coincidence (he still hasn't paid up). We now know that each of Ogg's groups $\Gamma_0(p)+$ is the genus-0 modular group for the function T_g, for some element $g \in \mathbb{M}$ of order p. Although we now realise why the Monster's primes must be a subset of Ogg's, probably there is no deep reason why Ogg's list couldn't have been longer.

The appeal of Monstrous Moonshine lies in its mysteriousness: it unexpectedly associates various special modular functions with the Monster, even though modular functions and elements of $\mathbb{M}$ are conceptually incommensurable. Now, 'understanding' something means to embed it naturally into a broader context. Why is the sky blue? Because of the way light scatters in gases. Why does light scatter in gases the way it does? Because of Maxwell's equations. In order to understand Monstrous Moonshine, to resolve the mystery, we should search for similar phenomena, and fit them all into the same story.

0.5 The Moonshine of E_8 and the Leech

McKay had also remarked in 1978 that similar numerology to (0.2.1) holds if $\mathbb{M}$ and $j(\tau)$ are replaced with the Lie group $E_8(\mathbb{C})$ and

$$j(\tau)^{\frac{1}{3}} = q^{-\frac{1}{3}}(1 + 248\,q + 4124\,q^2 + 34\,752\,q^3 + \cdots). \tag{0.5.1}$$

In particular, $4124 = 3875 + 248 + 1$ and $34\,752 = 30\,380 + 3875 + 2 \cdot 248 + 1$, where 248, 3875 and 30 380 are all dimensions of irreducible representations of $E_8(\mathbb{C})$. A *Lie group* is a manifold with compatible group structure; the groups of E_8 type play the same role in Lie theory that the Monster does for finite groups. Incidentally, $j^{\frac{1}{3}}$ is the Hauptmodul of the genus-0 group $\Gamma(3)$ (see (2.2.4a)).

A more elementary observation concerns the Leech lattice. A lattice is a discrete periodic set L in $\mathbb{R}^n$, and the Leech lattice Λ is a particularly special one in 24 dimensions. 196 560, the number of vectors in the Leech lattice with length-squared 4, is also close to 196 884: in fact,

$$196\,884 = 196\,560 + 324 \cdot 1, \tag{0.5.2a}$$

$$21\,493\,760 = 16\,773\,120 + 24 \cdot 196\,560 + 3200 \cdot 1, \tag{0.5.2b}$$

$$864\,299\,970 = 398\,034\,000 + 24 \cdot 16\,773\,120 + 324 \cdot 196\,560 + 25\,650 \cdot 1, \tag{0.5.2c}$$

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

where 16 773 120 and 398 034 000 are the numbers of length-squared 6- and 8-vectors in the Leech. This may not seem as convincing as (0.2.1), but the same equations hold for any of the 24-dimensional even self-dual lattices, apart from an extra term on the right sides corresponding to length-squared 2 vectors (there are none of these in the Leech).

What conceptually does the Monster, E_8 and the Leech lattice have to do with the j-function? Is there a common theory explaining this numerology? The answer is yes!

It isn't difficult to relate E_8 to the j-function. In the late 1960s, Victor Kac [**325**] and Robert Moody [**430**] independently (and for entirely different reasons) defined a new class of infinite-dimensional Lie algebras. A *Lie algebra* is a vector space with a bilinear vector-valued product that is both anti-commutative and anti-associative (Section 1.4.1). The familiar vector-product $u \times v$ in three dimensions defines a Lie algebra, called $\mathfrak{sl}_2$, and in fact this algebra generates all Kac–Moody algebras. Within a decade it was realised that the graded dimensions of representations of the *affine* Kac–Moody algebras are (vector-valued) modular functions for $\mathrm{SL}_2(\mathbb{Z})$ (Theorem 3.2.3).

Shortly after McKay's E_8 observation, Kac [**326**] and James Lepowsky [**373**] independently remarked that the unique level-1 highest-weight representation $L(\omega_0)$ of the affine Kac–Moody algebra ${E_8}^{(1)}$ has graded dimension $j(q)^{\frac{1}{3}}$. Since each homogeneous piece of any representation $L(\lambda)$ of the affine Kac–Moody algebra ${X_\ell}^{(1)}$ must carry a representation of the associated finite-dimensional Lie group $X_\ell(\mathbb{C})$, and the graded dimensions (multiplied by an appropriate power of q) of an affine algebra are modular functions for some $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$, this explained McKay's E_8 observation. His Monster observations took longer to clarify because so much of the mathematics needed was still to be developed.

Euler played with a function $t(x) := 1 + 2x + 2x^4 + 2x^9 + 2x^{16} + \cdots$, because it counts the ways a given number can be written as a sum of squares of integers. In his study of elliptic integrals, Jacobi (and Gauss before him) noticed that if we change variables by $x = e^{\pi \mathrm{i} \tau}$, then the resulting function $\theta_3(\tau) := 1 + 2e^{\pi \mathrm{i} \tau} + 2e^{4\pi \mathrm{i} \tau} + \cdots$ behaves nicely with respect to certain transformations of τ – we say today that Jacobi's theta function θ_3 is a modular form of weight $\frac{1}{2}$ for a certain index-3 subgroup of $\mathrm{SL}_2(\mathbb{Z})$. More generally, something similar holds when we replace $\mathbb{Z}$ with any other lattice L: the theta series

$$\Theta_L(\tau) := \sum_{n \in L} e^{\pi \mathrm{i} n \cdot n \tau}$$

is also a modular form, provided all length-squares $n \cdot n$ are rational. In particular, we obtain quite quickly that the theta series of the Leech lattice, divided by Ramanujan's modular form $\Delta(\tau)$, will equal $J(\tau) + 24$.

For both E_8 and the Leech, the j-function arises from a uniqueness property ($L(\omega_0)$ is the only 'level-1' ${E_8}^{(1)}$-module; the Leech lattice Λ is self-dual), together with the empirical observation that $\mathrm{SL}_2(\mathbb{Z})$ has few modular forms of small level. In these examples, the appearance of the j-function isn't as significant as that of modularity.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

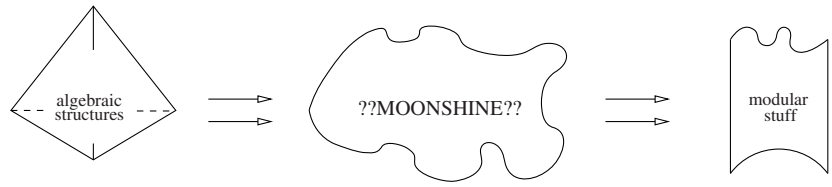

Fig. 0.1 Moonshine in its broader sense.

0.6 Moonshine beyond the Monster

We've known for many years that lattices (quadratic forms) and Kac–Moody algebras are related to modular forms and functions. But these observations, albeit now familiar, are also a little mysterious, we should confess. For instance, compare the non-obvious fact that $\theta_3(-1/\tau) = \sqrt{\frac{\tau}{\mathrm{i}}}\theta_3(\tau)$ with the trivial observation (0.1.6) that $G_k(-1/\tau) = \tau^k G_k(\tau)$ for the Eisenstein series G_k. The modularity of G_k is a special case of the elementary observation that $\mathrm{SL}_n(\mathbb{Z})$ parametrises the change-of-bases of n-dimensional lattices. The modularity of θ_3, on the other hand, begs a *conceptual* explanation (indeed, see the quote by Weil at the beginning of Section 2.4.2), even though its *logical* explanation (i.e. proof) is a quick calculation from, for example, the Poisson summation formula (Section 2.2.3). Moonshine really began with Jacobi and Gauss.

> *Moonshine should be regarded as a certain collection of related examples where algebraic structures have been associated with automorphic functions or forms.*

Grappling with that thought is the theme of our book. Chapters 1 to 6 could be (rather narrowly) regarded as supplying a context for Monstrous Moonshine, on which we focus in Chapter 7. From this larger perspective, illustrated in Figure 0.1, what is special about this single instance called *Monstrous* Moonshine is that the several associated modular functions are all of a special class (namely Hauptmoduls).

The first major step in the proof of Monstrous Moonshine was accomplished in the mid-1980s with the construction by Frenkel–Lepowsky–Meurman [**200**] of the Moonshine module $V^\natural$ and its interpretation by Richard Borcherds [**68**] as a *vertex operator algebra*. A vertex operator algebra (VOA) is an infinite-dimensional vector space with infinitely many heavily constrained vector-valued bilinear products (Chapter 5). It is a natural, though extremely intricate, extension of the notion of a Lie algebra. Any algebra $\mathcal{A}$ can be interpreted as an assignment of a linear map $\mathcal{A} \otimes \cdots \otimes \mathcal{A} \to \mathcal{A}$ to each binary tree; from this perspective a VOA $\mathcal{V}$ associates a linear map $\mathcal{V} \otimes \cdots \otimes \mathcal{V} \to \mathcal{V}$ with each 'inflated' binary tree, that is each sphere with discs removed.

In 1992 Borcherds [**72**] completed the proof of the original Monstrous Moonshine conjectures[4] by showing that the graded characters T_g of $V^\natural$ are indeed the Hauptmoduls identified by Conway and Norton, and hence that $V^\natural$ is indeed the desired representation

[4] As we see in Chapter 7, most Moonshine conjectures involving the Monster are still open.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

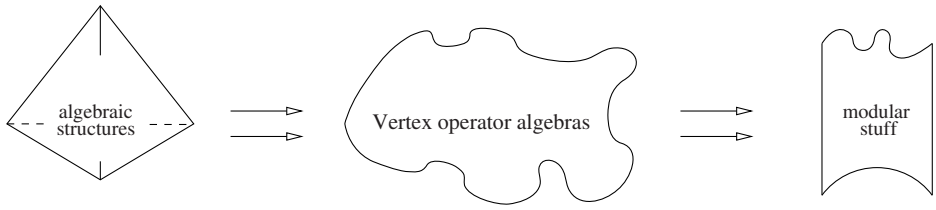

Fig. 0.2 The algebraic meaning of Moonshine.

V of $\mathbb{M}$ conjectured by McKay and Thompson. The explanation of Moonshine suggested by this picture is given in Figure 0.2. The algebraic structure typically arises as the symmetry group of the associated VOA – for example, that of $V^\natural$ is the Monster $\mathbb{M}$. By Zhu's Theorem (Theorem 5.3.8), the modular forms/functions appear as graded dimensions of the (possibly twisted) modules of the VOA. In particular, the answer this framework provides for what $\mathbb{M}$, E_8 and the Leech have to do with j is that they each correspond to a VOA with a single simple module; their relation to j is then an immediate corollary to the much more general Zhu's Theorem.

It must be emphasised that Figure 0.2 is primarily meant to address Moonshine in the broader sense of Figure 0.1, so certain special features of, for example, Monstrous Moonshine (in particular that all the T_g are Hauptmoduls) are more subtle and have to be treated by special arguments. These are quite fascinating by themselves, and are discussed in Chapter 7. Even so, Figure 0.2 provides a major clue:

> *If you're trying to understand a seemingly mysterious occurrence of the Monster, try replacing the word 'Monster' with its synonym 'the automorphism group of the vertex operator algebra $V^\natural$'.*

This places the Monster into a much richer algebraic context, with numerous connections with other areas of mathematics.

0.7 Physics and Moonshine

Moonshine is profoundly connected with physics (namely conformal field theory and string theory). String theory proposes that the elementary particles (electrons, photons, quarks, etc.) are vibrational modes on a string of length about 10^{-33} cm. These strings can interact only by splitting apart or joining together – as they evolve through time, these (classical) strings will trace out a surface called the *world-sheet*. Quantum field theory tells us that the quantum quantities of interest (amplitudes) can be perturbatively computed as weighted averages taken over spaces of these world-sheets. Conformally equivalent world-sheets should be identified, so we are led to interpret amplitudes as certain integrals over moduli spaces of surfaces. This approach to string theory leads to a conformally invariant quantum field theory on two-dimensional space-time, called *conformal field theory (CFT)*. The various modular forms and functions arising in Moonshine appear as integrands in some of these genus-1 ('1-loop') amplitudes: hence their modularity is manifest.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

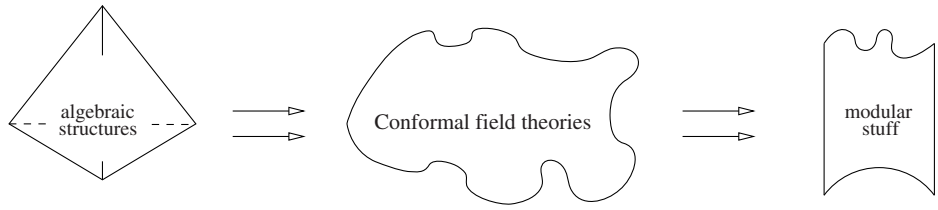

Fig. 0.3 The stringy picture of Moonshine.

Many aspects of Moonshine make complete sense within CFT, something which helps make the words of Freeman Dyson ring prophetic:

> I have a sneaking hope, a hope unsupported by any facts or any evidence, that sometime in the twenty-first century physicists will stumble upon the Monster group, built in some unsuspected way into the structure of the universe [**167**].

All that said, here we are, sometime in the twenty-first century, and alas the Monster still plays at best a peripheral role in physics. And some aspects of Moonshine (e.g. the Hauptmodul property) remain obscure in CFT. In any case, although this is primarily a mathematics book, we often sit in chairs warmed by physicists. In particular, CFT (or what is essentially the same thing, perturbative string theory[5]) is, at least in part, a machine for producing modular functions. Here, Figure 0.2 becomes Figure 0.3. More precisely, the algebraic structure is an underlying symmetry of the CFT, and its graded dimensions are the various modular functions. VOAs can be regarded as an algebraic abstraction of CFT, since they arise quite naturally by applying the Wightman axioms of quantum field theory to CFT. The lattice theta functions come from bosonic strings living on the torus $\mathbb{R}^n/L$. The affine Kac–Moody characters arise when the strings live on a Lie group. And the Monster is the symmetry of a string theory for a $\mathbb{Z}_2$-orbifold of free bosons compactified on the Leech lattice torus $\mathbb{R}^{24}/\Lambda$.

Physics reduces Moonshine to a duality between two different pictures of quantum field theory: the Hamiltonian one, which concretely gives us from representation theory the graded vector spaces, and another, due to Feynman, which manifestly gives us modularity. In particular, physics tells us that this modularity is a topological effect, and the group $\mathrm{SL}_2(\mathbb{Z})$ directly arises in its familiar role as the modular group of the torus.

Historically speaking, Figure 0.3 preceded and profoundly affected Figure 0.2. One reason the stringy picture is exciting is that the CFT machine in Figure 0.3 outputs much more than modular functions – it creates automorphic functions and forms for the various mapping class groups of surfaces with punctures. And all this is still poorly explored. We can thus expect more from Moonshine than Figure 0.2 alone suggests. On the other hand, once again Figure 0.3 can directly explain only the broader aspects of Moonshine.

[5] Curiously, although nonperturbative string theory should be *physically* more profound, it is the perturbative calculations that are most relevant to the *mathematics* of Moonshine.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

0.8 Braided #0: the meaning of Moonshine

In spite of the work of Borcherds and others, the special features of Monstrous Moonshine still beg questions. The full conceptual relationship between the Monster and Hauptmoduls (like j) arguably remains 'dimly lit', although much progress has been realised. This is a subject where it is much easier to speculate than to prove, and we are still awash in unresolved conjectures. But most important, we need a second independent proof of Monstrous Moonshine. In order to clarify the still murky significance of the Monster in Moonshine, we need to understand to what extent Monstrous Moonshine determines the Monster. More generally, we need to go beneath the algebraic explanation of Moonshine in order to find its more fundamental meaning, which is probably topological. Explaining something (Moonshine in this case) with something more complicated (CFT or VOAs here) cannot be the end of the story. Surely it is instead a beginning.

To Poincaré 125 years ago, modularity arose through the monodromy of differential equations. Remarkably, today CFT provides a similar explanation, although the relevant partial differential equations are much more complicated. The monodromy group here is the braid group $\mathcal{B}_3$, and the modular group $\mathrm{SL}_2(\mathbb{Z})$ arises as a homomorphic image.

Today we are taught to lift modular forms for $\mathrm{SL}_2(\mathbb{Z})$ to the space $L^2(\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R}))$, which carries a representation of the Lie group $\mathrm{SL}_2(\mathbb{R})$. However, $\mathrm{SL}_2(\mathbb{R})$ is not simply connected; its universal cover $\widetilde{\mathrm{SL}_2(\mathbb{R})}$ is a central extension by $\mathbb{Z}$, and the corresponding central extension of $\mathrm{SL}_2(\mathbb{Z})$ – the fundamental group of $\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ – is the braid group $\mathcal{B}_3$. By all rights, these central extensions should be more fundamental. Indeed, modular forms of fractional weight, such as the Dedekind eta, certainly see $\mathcal{B}_3$ more directly than they do $\mathrm{SL}_2(\mathbb{Z})$ (Section 2.4.3). Similar comments hold for other Γ – for example, the congruence subgroup $\Gamma(2)$ lifts to the pure braid group $\mathcal{P}_3$.

The best approach we know for relating the Monster and the Hauptmodul property is Norton's action of $\mathcal{B}_3$ on $G \times G$. This associates a genus-0 property with '6-transposition groups', which in turn points to a special role for $\mathbb{M}$, as the Monster is expected to be essentially the largest such group (Section 7.3.3). Incidentally, the number '6' arises here because the principal congruence subgroup $\Gamma(N)$ is genus 0 iff $N < 6$.

For these reasons and others we explore on the following pages, we expect a new proof for Moonshine to involve the braid group $\mathcal{B}_3$. The modular groups $\mathrm{SL}_2(\mathbb{Z})$ and $\mathrm{PSL}_2(\mathbb{Z})$ arise only indirectly as quotients. We also identify other promising places to look for alternate arguments for Moonshine – for example, the partial differential equations of CFT are built from the heat kernel, which has a long historical association with modularity.

0.9 The book

Borcherds' paper [**72**] and the resulting Fields medal close the opening chapter of the story of Moonshine. Now, 25 years after its formulation in [**111**], we are in a period of consolidation and synthesis, flames fanned I hope by this book.

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press

Most of us might liken much of our research to climbing a steep hill against a stiff breeze: every so often we stumble and roll to the bottom, but with persistence we eventually reach the summit and plant our flag amongst the others already there. And before our bruises fade and bones mend, we're off to the next hill. But perhaps research in its purest form is more like chasing squirrels. As soon as you spot one and leap towards it, it darts away, zigging and zagging, always just out of reach. If you're a little lucky, you might stick with it long enough to see it climb a tree. You'll never catch the damned squirrel, but chasing it will lead you to a tree. In mathematics, the trees are called theorems. The squirrels are those nagging little mysteries we write at the top of many sheets of paper. We never know where our question will take us, but if we stick with it, it'll lead us to a theorem. That I think is what research ideally is like. There is no higher example of this than Moonshine.

This book addresses the theory of the blob of Figure 0.1. We explore some of its versatility in Chapter 6, where we glimpse Moonshine orthogonal to the Monster. Like moonlight itself, Monstrous Moonshine is an indirect phenomenon. Just as in the theory of moonlight one must introduce the sun, so in the theory of Moonshine one must go well beyond the Monster. Much as a book discussing moonlight may include paragraphs on sunsets or comet tails, so do we discuss fusion rings, Galois actions and knot invariants. The following chapters use Moonshine (Monstrous and otherwise) as a happy excuse to take a rather winding little tour through modern mathematics and physics. If we offer more questions and suggestions than theorems and answers, at least that is in Moonshine's spirit.

This is not a textbook. The thought bobbing above my head like a balloon while writing was that the brain is driven by the *qualitative* – at the deepest level those are the only truths we seek and can absorb. I'm trying to share with the reader my understanding (such as it is) of several remarkable topics that fit loosely together under the motley banner *Moonshine*. I hope it fills a gap in the literature, by focusing more on the ideas and less on the technical minutiae, important though they are. But even if not, it was a pleasure to write, and I think that comes across on every page.

This book is philosophic and speculative, because Moonshine is. It is written for both physicists and mathematicians, because both subjects have contributed to the theory. Partly for this reason, this book differs from other mathematics books in the lack of formal arguments, and differs from other physics books in the lack of long formulae. Without doubt this will froth many mouths. Because the potential readership for this story is unusually diverse, I have tried to assume minimal formal background. Hence when you come to shockingly trivial passages or abrasively uninteresting tangents, please realise they weren't written for you.

In modern mathematics there is a strong tendency towards formulations of concepts that minimise the number and significance of arbitrary choices. This crispness tends to emphasise the naturality of the construction or definition, at the expense sometimes of accessibility. Our mathematics is more conceptual today – more beautiful perhaps – but the cost of less explicitness is the compartmentalism that curses our discipline. We have cut ourselves off not only from each other, but also from our past. In this book I've tried

to balance this asceticism with accessibility. Some things have surely been lost, but some perhaps have been gained.

The book endures some glaring and painful omissions, due mostly to fear of spousal reprisals were I to miss yet another deadline. I hope for a second edition. In it I would include a gentle introduction to geometric Langlands. I'd correct the total disregard here for all things supersymmetric – after all, most of the geometric impact of string theory involves supersymmetry. The mathematical treatment of CFT in Chapter 4 is sparser than I'd like. Section 5.4 was originally planned to include brief reviews of the chiral algebras of Beilinson–Drinfel'd [**48**] and the coordinate-free approach to VOAs developed in [**197**]. Cohomological issues arise in every chapter, where they are nonetheless quietly ignored. The lip-service paid to subfactors does no justice to their beautiful role in the theory.

I will probably be embarrassed five years from now as to what today I feel is *important*. But at worst I'll be surprised five years from now at what today I find *interesting*. The topics were selected based on my present interests. Other authors (and even me five years from now) would make different choices, but for that I won't apologise.

So let the chase begin. . .

https://doi.org/10.1017/9781009401548.001 Published online by Cambridge University Press