Seeing Galois Theory on Riemann Surfaces with Dessins d'Enfants

Will Dana

July 11, 2017

Contents

1 Introduction **1**

2 Riemann Surfaces and Coverings **3**
2.1 Riemann Surfaces 3
2.2 Morphisms and Ramification 5
2.3 Coverings 8
2.4 Algebraicity 15

3 Belyi's Theorem **21**
3.1 Galois Actions 22
3.2 (1) $\Longrightarrow$ (2) 23
3.3 An Alternative Criterion for (1) 27
3.4 Proving the Criterion 27
3.5 (2) $\Longrightarrow$ (1) 32
3.6 Aside: Obvious vs. Non-Obvious 32

4 Correspondences **33**
4.1 Object 1: Dessins d'Enfants 34
4.2 Object 2: Constellations and Permutations 39
4.3 Object 3: Thrice-Ramified Coverings 41
4.4 Moving Between Dessins and Permutations 42
4.5 Moving Between Permutations and Coverings 46

5 Conclusion **47**

1 Introduction

In 1984, Alexander Grothendieck (then retired, and beginning to withdraw from the mathematical community altogether) wrote, but did not publish, a paper entitled *Esquisse d'un Programme* ("Sketch of a Program"). The manuscript

was meant as an outline of a bold potential research problem connecting the Galois theory of the algebraic numbers, the moduli spaces of algebraic curves, and many other topics.

One inspiration for the work in the *Esquisse* was a theorem proved five years earlier by Belyi, the converse ($2 \Rightarrow 1$) of which Grothendieck had already known:

Theorem 1.1 (Belyi's Theorem)**.** *For a Riemann surface S, the following are equivalent:*

1. *S is isomorphic to an algebraic curve with coefficients in $\overline{\mathbb{Q}}$, the algebraic numbers.*
2. *There exists a holomorphic map from S to the projective line ramifying over at most 3 points.*

This unexpected equivalence between a somewhat number-theoretic statement and a statement of complex analysis suggests tempting connections between areas of mathematics which seem unrelated on the surface. In pursuing this, the *Esquisse* introduces different classes of objects which are equivalent to the covers considered by the theorem—most famously, as **dessins d'enfants**, or "children's drawings", bicolored graphs embedded on topological surfaces. The long-term hope is that, by studying this connection between graphs and algebraic numbers, one could gain an understanding of the absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ in combinatorial terms.

What was striking to Grothendieck, and remains striking today, is the relatively elementary nature of the objects and correspondences involved. One can describe the thrice-ramified coverings as graphs on surfaces, or pairs of permutations, objects which seem very simple by themselves, but come with different sets of tools that connect in unexpected ways. This simplicity made for a contrast with Grothendieck's earlier, more famous work, something he commented on in the introduction to the *Esquisse*:

> "Whereas in my research before 1970, my attention was systematically directed towards objects of maximal generality…here I was brought back, via objects so simple that a child learns them while playing, to the beginnings and origins of algebraic geometry, familiar to Riemann and his followers!" [6]

The thought that something basic on the surface could lend insight into deep questions of algebraic geometry and number theory has made the study of dessins d'enfants a small but active field to this day.

This paper proceeds in three parts. In the first part, we quickly outline the basic properties and definitions relevant to Riemann surfaces that we will need for the other two, with a particular focus on covering maps, ramification, and monodromy. In the second part, we prove Belyi's theorem. In the third part, we discuss the correspondences between maps of Riemann surfaces, dessins d'enfants, and pairs of permutations. Our primary sources in this endeavor

are Girondo and Gonzalez-Diez's *Introduction to Compact Riemann Surfaces and Dessins d'Enfants* [3] (for the first two parts) and Lando and Zvonkin's *Graphs on Surfaces and Their Applications* [8] (for the third). Future versions of this paper are anticipated to include a fourth part on the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on dessins d'enfants.

2 Riemann Surfaces and Coverings

The basic objects relevant to Belyi's theorem and dessins d'enfants are ramified coverings of Riemann surfaces, the definition and properties of which we recap here. A great deal of complex analysis goes into establishing these properties, and we will not be able to give this aspect the attention it deserves in this brief overview. We will omit or sketch most proofs, with the exception of the main theorem of Section 2.3.2, which is crucial to the rest of the paper.

We follow the introductions given in [3] and [11], occasionally referring to [10] for topological background.

2.1 Riemann Surfaces

Definition 2.1. *A **Riemann surface** is a connected topological space S, together with a covering by open sets U_i and homeomorphisms $\varphi_i : U_i \to \varphi_i(U_i) \subset \mathbb{C}$ (the pairs of which are called **charts**) such that the **transition maps***

$$\varphi_i \circ \varphi_j^{-1} : \varphi_j(U_i \cap U_j) \to \varphi_i(U_i \cap U_j)$$

are holomorphic.

The motivation behind this definition is an interest in defining what it means for functions on topological surfaces to be holomorphic. The charts allow us to translate the question back to the complex plane. If S is a Riemann surface with charts (U_i, φ_i), and $p \in U_i$, then we say a functon $f : S \to \mathbb{C}$ is holomorphic at p if $f \circ \varphi_i^{-1} : \varphi(U_i) \to \mathbb{C}$, the **coordinate representation** of f, is. The condition on the transition maps is then necessary to show that this is well-defined. If (U_j, φ_j) is another chart with $p \in U_j$, then since $f \circ \varphi_j^{-1} = (f \circ \varphi_i^{-1}) \circ (\varphi_i \circ \varphi_j^{-1})$ is a composition of holomorphic functions, it is also holomorphic.

Disregarding the holomorphy condition and interpreting $\mathbb{C}$ as $\mathbb{R}^2$, the existence of the charts makes S into a topological surface (2-manifold). Disregarding the complex aspect of the holomorphy condition shows that S also has a smooth structure. In this latter perspective, one also sees that the underlying surface is naturally orientable: one way of specifying an orientation is by choosing charts such that the Jacobian determinants of the transition maps $\varphi_i \circ \varphi_j^{-1}$ are positive, and if the transition maps are analytic this follows straightforwardly from the Cauchy-Riemann equations. Intuitively, the complex plane has a notion of counterclockwise rotation built into it, given by multiplication by i, and this translates into an orientation on a Riemann surface.

Now we give two important examples of Riemann surfaces.

Example: Affine Plane Curves Let $f(x, y)$ be a polynomial in two variables, such that f and its partial derivatives f_x, f_y are never simultaneously 0. Then we can put a Riemann surface structure on the zero locus $S = \{(x, y) \in \mathbb{C}^2 : f(x, y) = 0\}$ as follows.

Given $(x_0, y_0) \in S$, by assumption either $f_x(x_0, y_0)$ or $f_y(x_0, y_0)$ is nonzero, without loss of generality the latter. By the Implicit Function Theorem, there exists a holomorphic function g defined in a neighborhood of x_0 and a neighborhood U of (x_0, y_0) such that $U \cap S$ consists of exactly the points $(x, g(x))$. Then we can simply define a chart φ on this neighborhood by projection onto the first coordinate. To see why these charts are compatible, suppose that there is another neighborhood on which the points of S are given by $(h(y), y)$ for some holomorphic function h, and a chart ψ is defined by projection onto the second coordinate. Then

$$\varphi \circ \psi^{-1}(y) = \varphi(h(y), y) = h(y)$$
$$\psi \circ \varphi^{-1}(x) = \psi(x, g(x)) = g(x)$$

and these are holomorphic.

The affine plane curves just constructed are not compact. For the remainder of the paper, we will primarily be interested in compact Riemann surfaces, which are outstandingly nice in a few ways. (In particular, Riemann surfaces will be assumed compact unless stated otherwise.) For example, we can refer to the classification of compact orientable surfaces [10, Theorem 10.22] to associate to each Riemann surface the **genus** of its underlying topological surface. With this in mind, we now look at the simplest compact Riemann surface.

Example: The Projective Line Starting with the topological space $\mathbb{C}^2 \backslash \{(0, 0)\}$, we identify points which can be obtained from each other by scaling:

$$(z, w) \sim (\lambda z, \lambda w), \quad \lambda \in \mathbb{C} \backslash \{0\}$$

We consider the quotient space under this identification, which we call $\mathbb{P}^1$, the **one-dimensional complex projective space**. To distinguish it from $\mathbb{C}^2$, we denote the equivalence class of (z, w) by $[z : w]$. We give $\mathbb{P}^1$ the structure of a Riemann surface as follows.

We cover $\mathbb{P}^1$ by the open sets

$$U_0 = \{[z : w] \mid z \neq 0\}$$
$$U_1 = \{[z : w] \mid w \neq 0\}$$

and then define the maps $\varphi_i : U_i \to \mathbb{C}$ by

$$\varphi_0[z : w] = w/z$$
$$\varphi_1[z : w] = z/w$$

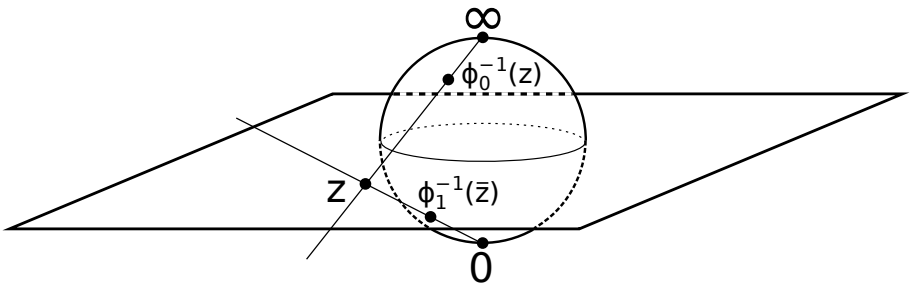


Figure 1: The Riemann sphere, with its two charts φ_0 and φ_1.

It is straightforward to verify that these are well-defined homeomorphisms on projective space, and we have $\varphi_0 \circ \varphi_1^{-1}(z) = \varphi_0[z:1] = 1/z$ and $\varphi_1 \circ \varphi_0^{-1}(w) = 1/w$, which are holomorphic.

$\mathbb{P}^1$ admits a variety of descriptions. First, we note that any point $[z:w]$ can be placed in a unique canonical form as either $[z/w:1]$, if $w \neq 0$, or $[1:0]$, if $w = 0$. We can identify the points $[z:1]$ with $\mathbb{C}$, and refer to the point $[1:0]$ as ∞, and we will often use the language of $\mathbb{C} \cup \{\infty\}$ to talk about $\mathbb{P}^1$. In this notation, the charts we gave for $\mathbb{P}^1$ become:

$$U_0 = \mathbb{C}, \quad \varphi_0(z) = z$$
$$U_1 = \mathbb{C} \cup \{\infty\} \backslash \{0\}, \quad \varphi_1(z) = \begin{cases} 1/z & z \neq \infty \\ 0 & z = \infty \end{cases}$$

Another Riemann surface isomorphic to $\mathbb{P}^1$ (in the sense discussed in the next section) is the **Riemann sphere**, consisting of the sphere with two charts given by stereographic projection from the north and south poles (the latter composed with a complex conjugation). This is illustrated in Figure 1. Again, this gives a natural notion of a point at ∞: stereographic projection from the north pole gives a bijection between the the sphere without its north pole and all of $\mathbb{C}$, so we can refer to the north pole as ∞.

We will generally call this surface $\mathbb{P}^1$ or $\mathbb{C} \cup \{\infty\}$; however, one important fact to take away from this discussion is that $\mathbb{P}^1$ has genus 0. We may on occasion refer to it as "the sphere".

2.2 Morphisms and Ramification

Just as we defined holomorphic functions on Riemann surfaces above, we now define holomorphic maps between Riemann surfaces, which are the morphisms in the relevant category.

Definition 2.2. *Let S, S' be Riemann surfaces with systems of charts $\{(U_i, \varphi_i)\}$ and $\{(U_i', \varphi_i')\}$, respectively. Then a continuous function $f : S \to S'$ is **holomorphic at** p if, for charts (U_i, φ_i) and (U_j', φ_j') such that $p \in U_i$ and $f(p) \in U_j'$,*

$$\varphi_j' \circ f \circ \varphi_i^{-1}$$

is holomorphic.

Again, this is well-defined by the compatibility condition on the charts. In fact, we can carry over a bit more from the corresponding concept on the complex plane. Recall that, if $f : \mathbb{C} \to \mathbb{C}$ is holomorphic and $f(z_0) = 0$, then the **order of the zero** at z_0 is the smallest m such that the coefficient of $(z - z_0)^m$ in the Taylor series of f is nonzero, or equivalently the smallest m such that $f^{(m)}(z_0) \neq 0$.

More generally, we define the **multiplicity** of f at any point z_0 to be the order of the zero at $f(z) - f(z_0)$, or equivalently the smallest $m \geq 1$ such that $f^{(m)}(z_0) \neq 0$. Just as with order, the multiplicity of a composition $f \circ g$ is the product of the multiplicities of f and g at the points involved.

Definition 2.3. *Let $f : S \to S'$ be a holomorphic map of Riemann surfaces. The **multiplicity of** f **at** p, denoted $\mathrm{mult}_p(f)$, is (in the notation of Definition 2.2) the multiplicity of $\varphi'_j \circ f \circ \varphi_i^{-1}$ at $\varphi_i(p)$.*

As above, switching to a different choice of chart means composing the coordinate representation with some $\varphi_i \circ \varphi_k^{-1}$ or $\varphi'_k \circ {\varphi'_j}^{-1}$. These are one-to-one holomorphic maps, and so by a basic result of complex analysis, their derivatives must be nonzero, and they have multiplicity 1. Then composing with these maps does not change the multiplicity of f, which is thus well-defined.

Definition 2.4. *A point p for which $\mathrm{mult}_p(f) > 1$ is a **ramification point** of f. The value $f(p)$ is a **branch value**. We say that f **ramifies over** $f(p)$.*

We will not prove the following result, but it is an important demonstration of the appropriateness of the term "multiplicity".

Theorem 2.5 ([11], Proposition 4.8)**.** *Let $f : S \to S'$ be a holomorphic map of Riemann surfaces, and let $q \in S'$. The number*

$$\sum_{p \in f^{-1}(q)} \mathrm{mult}_p(f)$$

is independent of q.

This quantity is called the degree of f, or $\deg(f)$. Over any point that's not a branch value (which is "most" points, as branch values are isolated), f is $\deg(f)$-to-1. Branch values are the points where the fiber is smaller than this.

We note that $\deg(f)$ is finite, which we sketch here. Given a non-branch value $y \in Y$, the points of the fiber $f^{-1}(\{y\})$ must be isolated, which follows from the corresponding fact for nonconstant holomorphic functions on $\mathbb{C}$. If there were infinitely many, we could use disjoint open sets around all of them together with the complement of the fiber to construct an open cover of X with no finite subcover, contradicting compactness.

Finally, we consider meromorphic functions on a Riemann surface. Just as with the other concepts we've translated over from complex analysis, a function f is meromorphic if, for every chart (U_i, φ_i), the composition $f \circ \varphi_i^{-1}$ is

meromorphic. There is an alternative characterization of meromorphic functions, based on our description of $\mathbb{P}^1$ as $\mathbb{C} \cup \{\infty\}$.

Proposition 2.6. *A function $f : S \to \mathbb{C} \cup \{\infty\}$ is meromorphic if and only if the corresponding function $S \to \mathbb{P}^1$ is holomorphic.*

Proof. We have that $f : S \to \mathbb{C} \cup \{\infty\}$ (resp. $f : S \to \mathbb{P}^1$) is meromorphic (resp. holomorphic) if, for every chart (U_i, φ_i) of S, the compositions $f \circ \varphi_i^{-1}$ are as functions on subsets of $\mathbb{C}$. Thus it suffices to treat the case of subsets of $\mathbb{C}$.

In this case, a basic result of complex analysis is that a function f is meromorphic if and only if it is a quotient of two holomorphic functions $g(z)/h(z)$. Under the correspondence between $\mathbb{C} \cup \{\infty\}$ and $\mathbb{P}^1$, this corresponds to the function to $\mathbb{P}^1$ given by

$$z \mapsto [g(z) : h(z)]$$

as this is ∞ exactly when $h(z) = 0$. (For this to be well-defined at a point z_0, one of $g(z_0), h(z_0)$ must be nonzero, but we can ensure this in a neighborhood of z_0 by dividing $g(z)$ and $h(z)$ by the highest power of $z - z_0$ dividing both of them.) Then for a general g and h, the function $z \mapsto [g(z) : h(z)]$ is holomorphic exactly when $g(z)/h(z)$ and $h(z)/g(z)$ are holomorphic on their domains of definition. However, this is equivalent to $g(z)/h(z)$ being meromorphic.

□

For this reason we will refer to "holomorphic maps to $\mathbb{P}^1$" and "meromorphic functions" interchangeably.

For future reference, we note two properties of meromorphic functions that are easy to transfer from the complex plane to Riemann surfaces. Just as how we defined multiplicity above, if f is a meromorphic function and p is a point on a Riemann surface, we define the order $\mathrm{ord}_p(f)$ of f to be that of its coordinate representation $f \circ \varphi_i^{-1}$ at $\varphi_i(p)$.

Additionally, recall that the meromorphic functions on a subset of $\mathbb{C}$ form a field, based on the rule that the multiplicative inverse of a function has poles at its zeroes and vice versa. Given a Riemann surface S, we can similarly talk about its field of meromorphic functions, which we denote $\mathcal{M}(S)$.

2.2.1 Ramification of Maps $\mathbb{P}^1 \to \mathbb{P}^1$

Here we interpret what ramification means in the specific case of meromorphic functions on $\mathbb{P}^1$. In addition to providing a simple example of the definitions, this particular case will be useful in our proof of Belyi's theorem and in working out some examples of dessins d'enfants. We use the description of $\mathbb{P}^1$ as $\mathbb{C} \cup \{\infty\}$, and the corresponding charts, given above.

So suppose $f(p) = w$. If $p, w \in \mathbb{C}$, $\mathrm{mult}_p(f)$ reduces to multiplicity in the ordinary sense on the complex plane. If $p \in \mathbb{C}$ but $w = \infty$, then $\mathrm{mult}_p(f)$ is instead given by the multiplicity of $1/f$ at p. Finally, if $p = \infty$, then $\mathrm{mult}_\infty(f)$ is given by the multiplicity of $f(1/z)$ at $z = 0$.

It is a standard result of complex analysis [11, Theorem II.2.1] that the meromorphic functions on $\mathbb{C} \cup \{\infty\}$ are precisely the rational functions. Knowing

this, we can interpret the above conditions more precisely. Suppose $f(z) = g(z)/h(z)$, where g and h are relatively prime polynomials. If $f(p) = 0$, then $\mathrm{mult}_p(f)$ is the largest power of $(z - p)$ dividing g, just as with polynomials; thus if $f(p) = \infty$, then $\mathrm{mult}_p(f)$ is the highest power of $z - p$ dividing h.

To find the multiplicity of f at ∞, let d_g and d_h be the degrees of g and h, respectively, and let d be the larger of the two. Then

$$f(1/z) = \frac{g(1/z)}{h(1/z)} = \frac{z^d g(1/z)}{z^d h(1/z)}$$

If $d_h > d_g$, then this has a zero of order $d_h - d_g$ at 0; if $d_g > d_h$, then it has a pole of order $d_g - d_h$. Thus in these cases $\mathrm{mult}_\infty(f) = |d_g - d_h|$. The case $d_g = d_h$ must be treated separately.

Finally, we single out the automorphisms of $\mathbb{P}^1$; these are given by the **Möbius transformations**, or linear fractional transformations, rational functions of the form

$$\frac{az + b}{cz + d}$$

with $ad - bc \neq 0$. Importantly, the group of Möbius transformations is 3-transitive: for any triples of points z_1, z_2, z_3 and w_1, w_2, w_3, there is a Möbius transformation sending z_i to w_i. This will allow us, when talking about distinguished points on the Riemann sphere, to put 3 of them wherever we want without loss of generality, by applying a suitable automorphism.

2.3 Coverings

The fundamental objects of study in the theory of Belyi functions are not Riemann surfaces in isolation, but maps from these surfaces to the Riemann sphere. The niceness of holomorphic maps allows us to approach these maps using the theory of covering spaces, which we review here. (We refer the reader to Chapter 11 of [10] for further details.)

Definition 2.7. *Let X and Y be connected manifolds*[1]*. A **covering map** is a continuous surjection $q : X \to Y$ such that every point $y \in Y$ has a neighborhood U with the following property: $q^{-1}(U)$ is a disjoint union of sets U_α such that q maps each U_α homeomorphically onto U. Such a neighborhood is called **evenly covered**. X is the **covering space** and Y is the **base space**.*

One consequence of this definition is that all the fibers $q^{-1}(y)$ have the same size [10, Proposition 11.11]. In analogy with Theorem 2.5, we call this the **degree** of the covering.

[1]This definition can be made for general topological spaces satisfying certain technical connectivity conditions, but we have no reason to consider this here.

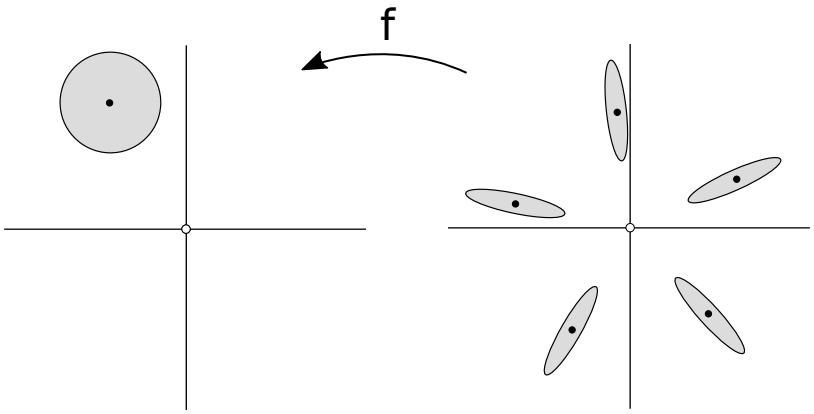


Figure 2: The nth power map $f : \mathbb{C}\backslash\{0\} \to \mathbb{C}\backslash\{0\}$ sending $z \mapsto z^n$, seen here with $n = 5$, is a covering map.

Example. Consider the function $f : \mathbb{C}\backslash\{0\} \to \mathbb{C}\backslash\{0\}$ given by $f(z) = z^n$. Then the inverse image of a sufficiently small disk around the point falls apart into n components evenly spaced around the origin, and f is a homeomorphism when restricted to each component. See Figure 2. Thus f is a covering of degree n.

However, if we extend the domain to consider the nth power map $f : \mathbb{C} \to \mathbb{C}$, it is not a covering map. If it were, then there would be a neighborhood U of 0 satisfying the definition of covering map. Since every component of $f^{-1}(U)$ must contain a point mapping to 0, and $f^{-1}(\{0\}) = \{0\}$, $f^{-1}(U)$ can only have one component. However, f is not 1-to-1 on any disk around 0, so it cannot restrict to a homeomorphism on this component.

Covering maps interact nicely with Riemann surfaces, in that we can "pull back" Riemann surface structures along them.

Lemma 2.8. *Let $q : X \to Y$ be a covering map, and suppose Y has the structure of a (possibly noncompact) Riemann surface. Then there is a unique way to give X a (possibly noncompact) Riemann surface structure such that q is holomorphic.*

Here "unique" means that, given any other collection of charts on X making q a holomorphic map, they are compatible with the charts implied by the lemma, in the sense that the transition maps between charts from the two collections are holomorphic.

Proof. Let $y \in Y$, and choose an evenly covered neighborhood $U_i \ni y$ small enough to lie within the domain of a chart $\varphi_i : U_i \to \varphi_i(U_i)$. Then for each of the components $U_{i,\alpha}$ of $f^{-1}(U_i)$, define $\varphi_{i,\alpha} : U_{i,\alpha} \to \varphi_i(U_i)$ by $\varphi_{i,\alpha} = \varphi_i \circ f$. By evenly-coveredness, this is still a homeomorphism, and the maps $\varphi_{j,\beta} \circ \varphi_{i,\alpha}^{-1} = \varphi_j \circ q \circ q^{-1} \circ \varphi_i = \varphi_j \circ \varphi_i$ are still holomorphic.

Then since $\varphi_j \circ q \circ \varphi_{i,\alpha}^{-1} = \varphi_j \circ q \circ q^{-1} \circ \varphi_i^{-1} = \varphi_j \circ \varphi_i^{-1}$ is holomorphic, q is holomorphic. The uniqueness is equally straightforward: if (V_i, ψ_i) is another collection of charts on X making f holomorphic, then we have $\varphi_{i,\alpha} \circ \psi_j^{-1} = \varphi_i \circ q \circ \psi_j^{-1}$, which is holomorphic by definition, and so the charts are compatible. □

Not all maps between Riemann surfaces are covering maps, but the two concepts turn out to be pretty close. We saw above that the obstruction to our covering map $z \mapsto z^n$ extending to all of $\mathbb{C}$ is the ramification at 0, where the map is 1-to-1 rather than n-to-1. This turns out to be a prototype for the behavior of all other holomorphic maps.

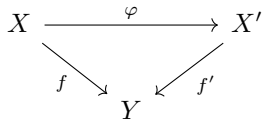
Theorem 2.9. *Let $f : X \to Y$ be a holomorphic map of Riemann surface, and let $\{y_1, \ldots, y_n\} \subset Y$ be its branch values. Let $Y^* = Y \backslash \{y_1, \ldots, y_n\}$ and $X^* = X \backslash f^{-1}(\{y_1, \ldots, y_n\}$. Then the restriction $f : X^* \to Y^*$ is a covering map.*

Proof. Sketch: Given a point $y \in Y^*$, it has finitely many preimages $f^{-1}(\{y\}) = \{x_1, \ldots, x_d\}$. At each x_i, the coordinate representation of f has nonzero derivative, so by the inverse function theorem x_i has a neighborhood on which f is a homeomorphism. Taking these neighborhoods to be small enough that they are disjoint, f then maps each one onto a neighborhood of y. The intersection of these gives the neighborhood of y required by the definition of covering map. □

So while covering maps are quite special in the general topological case, any morphism of Riemann surfaces can be made into a covering map by removing the points of ramification. For this reason, arbitrary morphisms of Riemann surfaces are sometimes called **ramified coverings**.

Since we want to consider ramified coverings as objects in themselves, we want a notion of morphisms between them.

Definition 2.10. *Let $f : X \to Y$ and $f' : X' \to Y$ be ramified coverings of a surface Y. A morphism $\varphi : f \to f'$ is a morphism of Riemann surfaces $\varphi : X \to X'$ such that $f' \circ \varphi = f$, that is, such that this diagram commutes:*



Having laid out the basic definitions, we move on to the most important property of (unramified) covering maps for our purposes: the **path lifting property** [10, Corollary 11.14]. We recall that, given points $a, b \in Y$, a **path** between a and b is a continuous map $\gamma : [0, 1] \to Y$ with $\gamma(0) = a$ and $\gamma(1) = b$.

Proposition 2.11 (Path Lifting Property)**.** *Suppose $q : X \to Y$ is a covering map, and $\gamma : [0, 1] \to Y$ is a path in Y between a and b. Let $\widetilde{a} \in q^{-1}(\{a\})$. Then there is a unique path $\widetilde{\gamma} : [0, 1] \to X$ such that $q \circ \widetilde{\gamma} = \gamma$ and $\widetilde{\gamma}(0) = \widetilde{a}$.*

In other words, any path in Y traces out a corresponding path in X, and once we pick the starting point $\widetilde{a}$ of that path, it is uniquely determined.

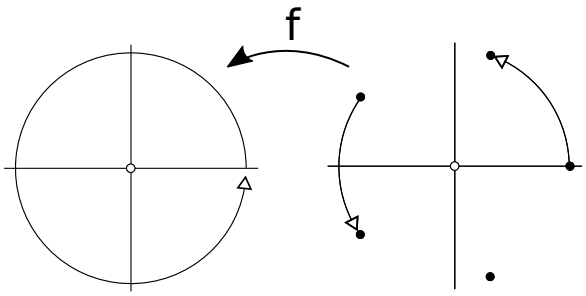


Figure 3: Lifting the circular loop around the origin along the covering map f sending $z \mapsto z^5$, with a couple of different starting points.

Example. Returning to the covering map $f : \mathbb{C}\backslash\{0\} \to \mathbb{C}\backslash\{0\}$, $z \mapsto z^n$ discussed above, we can consider the path $\gamma : [0,1] \to \mathbb{C}\backslash\{0\}$ given by $\gamma(t) = e^{2\pi it}$, which loops from 1 back to 1 along the unit circle. Then the points of the fiber $f^{-1}(\{1\})$ are the nth roots of unity. If we pick one of them, say ω, the lift of γ which starts at ω is given by $\widetilde{\gamma}(t) = \omega e^{2\pi it/n}$, and goes from ω to the next root of unity in counterclockwise order. This is illustrated in Figure 3.

2.3.1 Monodromy

After we form a covering map $f : X^* \to Y^*$ by removing all the points of ramification of a morphism, we can still get information about the ramification by using the path lifting property to define the monodromy group of the covering.

Recall that the **fundamental group** $\pi_1(Y^*, y)$ is the collection of homotopy equivalence classes of loops based at y (paths from y to itself). The product of two loops (which becomes a well-defined group operation on equivalence classes) is the loop obtained by following one loop followed by the other.

The basic idea of monodromy stems from lifting loops in the base space to paths in the covering space. While the loop goes from a point y back to itself, its lift may end up at a different preimage of y from the one it started at. The monodromy of the covering captures this in the form of a right action of $\pi_1(Y^*, y)$ on the fiber $f^{-1}(\{y\})$.

More precisely, given a path class $[\gamma] \in \pi_1(Y^*, y)$ with representative γ, and a point $\widetilde{y} \in f^{-1}(\{y\})$, the path lifting property tells us that there is a unique lift of γ, $\widetilde{\gamma}$, with $\widetilde{\gamma}(0) = \widetilde{y}$. Then we define $\widetilde{y} \cdot [\gamma] := \widetilde{\gamma}(1)$.

Of course, we need this to be well-defined, independent of the representative we choose for $[\gamma]$. This follows from the homotopy lifting property [10, Theorem 11.13].

This action is transitive: given two points $\widetilde{y}_1, \widetilde{y}_2 \in f^{-1}(y)$, we can produce a path $\widetilde{\gamma}$ between them, so that $f \circ \widetilde{\gamma}$ is a loop in Y^* which lifts to $\widetilde{\gamma}$, and $\widetilde{y}_1 \cdot [f \circ \widetilde{\gamma}] = \widetilde{y}_2$.

Example. Once more, we return to the map $f : \mathbb{C}\backslash\{0\} \to \mathbb{C}\backslash\{0\}, z \mapsto z^n$. It is a standard fact ([10, Corollary 8.10]) that $\pi_1(\mathbb{C}\backslash\{0\}, 1) \cong \mathbb{Z}$, with generator given by the class of the loop $\gamma(t) = e^{2\pi i t}$. We saw above that the lift of this path starting at a root of unity travels to the next root of unity counterclockwise. Thus the monodromy action of γ cyclically permutes the roots of unity in the fiber over 1.

As we've defined it, the monodromy of a covering depends on the choice of basepoint y in a couple of ways, but we can avoid this to some extent. If we label the points of $f^{-1}(\{y\})$ with $\{1, \ldots, d\}$, the monodromy action then takes the form of a homomorphism into the symmetric group $\pi_1(Y^*, y) \to S_d$.

Of course, this homomorphism will depend on our choice of labels as well as the choice of basepoint. We sidestep the former issue by noting that choosing the labels $\{1, \ldots, d\}$ differently just has the effect of conjugating by the permutation corresponding to the relabeling. So, up to conjugation, the homomorphism is independent of our labeling.

In the latter case, suppose $y' \in Y^*$ is a different basepoint. Recall that choosing a path p from y to y' gives a canononical isomorphism $\pi_1(Y^*, y) \to \pi_1(Y^*, y')$ sending $[\gamma]$ to the concatenated path $[p]^{-1}[\gamma][p]$. We can also get a bijection between the fibers over y and y' in the same way as we defined monodromy: for each $x \in f^{-1}(y)$, the path p lifts to a unique path from x to some $x' \in f^{-1}(y')$, and we denote $x' = x \cdot [p]$. Then one can check that these actions are compatible in the sense that $(x \cdot [\gamma]) \cdot [p] = (x \cdot [p]) \cdot [p]^{-1}[\gamma][p]$. Transferring the labels from $f^{-1}(y)$ to $f^{-1}(y')$ by the bijection defined by p gives a homomorphism $\pi_1(Y^*, y') \to S_d$ making the following diagram commute:

$$\begin{array}{ccc} \pi_1(Y^*, y) & \xrightarrow{[\gamma] \mapsto [p]^{-1}[\gamma][p]} & \pi_1(Y^*, y') \\ & \searrow \quad \swarrow & \\ & S_d & \end{array}$$

Now, one could use non-homotopic paths p and p' to define different isomorphisms $\pi_1(Y^*, y) \to \pi_1(Y^*, y')$. However, note that we can compare the images of a path $[\gamma]$ under the two isomorphisms by

$$[p]^{-1}[\gamma][p] = ([p]^{-1}[p'])([p']^{-1}[\gamma][p'])([p]^{-1}[p'])^{-1}.$$

So the two homomorphisms $\pi_1(Y^*, y') \to S_d$ obtained from the two paths differ only by conjugation by the loop $[p]^{-1}[p']$, which again corresponds to a relabeling of the fiber.

The outcome of this discussion is thus: when we talk about the monodromy of a covering of degree d, we refer to a homomorphism $\pi_1(Y^*, y) \to S_d$, where we identify homomorphisms which differ by conjugation in S_d. If we choose a different basepoint for our fundamental group, the resulting monodromy is related by a canonical isomorphism which, once the above identification is

made, is unique. Thus we will talk freely about the "monodromy of a covering" without specifying a base point or labeling of the fiber.

We refer to the image of $\pi_1(Y^*, y)$ in S_d as the **monodromy group** of the covering, bearing in mind that it is a conjugacy class rather than a well-defined subgroup.

Monodromy and the fundamental group give us powerful tools for classifying the covers of a given space. The culmination of this is the following theorem:

Theorem 2.12 ([10, Theorem 12.18])**.** *For fixed Y^* and y_0, there is a one-to-one correspondence between isomorphism classes of unramified coverings of Y^* and conjugacy classes in $\pi_1(Y^*, y_0)$. A covering $f : X^* \to Y^*$ corresponds to the conjugacy class of the stabilizer of any point in the fiber $f^{-1}(y_0)$ under the monodromy action. The degree of the covering equals the index of the subgroup.*

The last statement follows from the first by the orbit-stabilizer formula. Because the monodromy action is transitive action on $\deg(f)$ points, the stabilizer of any of these points has index $\deg(f)$. Implicit also in the theorem is the statement that all of the stabilizers are conjugate, which also follows from transitivity.

In fact, the stabilizer of a transitive group action is enough to determine it [10, Proposition 11.26b]. Thus the theorem can be rephrased as a correspondence between unramified coverings and transitive actions of $\pi_1(Y^*, y_0)$.

Example. We return to the example of a covering map that we have been considering all along, with a slight variation. Letting $D \subset \mathbb{C}$ be the unit disk, $f : D\backslash\{0\} \to D\backslash\{0\}$, $z \mapsto z^n$ is still a covering map of degree n. This is a covering of degree n. On the other hand, $\pi_1(D\backslash\{0\}, z_0) \cong \mathbb{Z}$, which has one subgroup of each integer index, and so the coverings we have found are the only ones of finite degree, up to isomorphism.

2.3.2 Ramified Covers with Prescribed Monodromy

The classification theorem above tells us that we can construct unramified coverings of a punctured Riemann surface with any transitive action as their monodromy action. In fact, we can extend this to a statement about ramified coverings of Riemann surfaces by carefully plugging holes in the base space with branch values.

Theorem 2.13. *Let Y be a Riemann surface, $R \subset Y$ a finite set, $y_0 \in Y\backslash R$, and $\mu : \pi_1(Y\backslash R, y_0) \to S_d$ a group homomorphism with transitive image. Then there exists a compact Riemann surface X and ramified covering $f : X \to Y$ of degree d, unique up to isomorphism of coverings, such that the monodromy of f is conjugate to μ.*

Proof. Let $Y^* = Y\backslash R$. By Theorem 2.12, there is a unique topological covering $f^* : X^* \to Y^*$ corresponding to the action defined by μ (via its stabilizer). Then

this covering already has the monodromy we want. Additionally, by Lemma 2.8, X^* can be given a unique (noncompact) Riemann surface structure making f^* holomorphic.

It remains to complete X^* to a Riemann surface X and complete f^* to a morphism $f : X \to Y$. A priori, "plugging the holes" in X^* is difficult, because we don't know what sort of "holes" we should be looking for, knowing only that X^* exists through the classification theorem. However, the classification of covers of the punctured disk $D\backslash\{0\}$ comes to our rescue.

At a point $y \in R$, we can choose a chart (U, φ) such that $\varphi(U) = D$ is the unit disk centered at $\varphi(y) = 0$ and U is sufficiently small that it contains no other points of R. Then ${f^*}^{-1}(U)$ splits into open connected components $V_1, \ldots, V_k$. We claim that each of the restrictions $f^*|_{V_j} : V_j \to U$ is a covering map.

First, we show that $f^*|_{V_j}$ is surjective. For any $y \in U$, we can choose a connected evenly covered neighborhood $N \ni y$, such that ${f^*}^{-1}(N)$ breaks into components homeomorphic to N. Then each of these components is also connected, and must be contained in one of the $V_{j'}$. If one of them is contained in V_j, then $N \subset f^*(V_j)$, while if none of them are, then N and $f^*(V_j)$ are disjoint. From this we can conclude that $f^*(V_j)$ is both open and closed in U; since U is connected, this means $f^*(V_j)$ is all of U. Additionally, since we have shown that each component of ${f^*}^{-1}(N)$ is entirely contained in some $V_{j'}$, N is still evenly covered when we restrict our attention to $(f^*|_{V_j})^{-1}(N)$.

Thus $f^*|_{V_j} : V_j \to U \cong D\backslash\{0\}$ is a covering of finite degree. By the classification of coverings of $D\backslash\{0\}$ worked out in the above example, there must exist an isomorphism ψ^* between this covering and one of the nth-power maps, making this diagram commute:

$$\begin{array}{ccc} V_j & \overset{\exists\psi^*}{\dashrightarrow} & D\backslash\{0\} \\ \downarrow{\scriptstyle f^*} & & \downarrow{\scriptstyle z\mapsto z^n} \\ U & \xrightarrow{\varphi} & D\backslash\{0\} \end{array}$$

Knowing that each of the components V_j is homeomorphic to a punctured disk gives us a natural way to fill holes: we add new points corresponding to the centers of the disks. For each V_j, we add a point x_j to X^*. To extend the Riemann surface structure to account for these points, we define a chart about x_j by $(V_j \cup \{x_j\}, \psi)$, where

$$\psi(x) = \begin{cases} \psi^*(x) & x \neq x_j \\ 0 & x = x_j \end{cases}$$

We must show this chart is compatible with the ones already existing on X^*; by definition, this amounts to showing that ψ^* is holomorphic. Let $x \in V_j$; then from the commutative diagram above, we know that $\psi^*(x)^n = \varphi \circ f^*(x)$. Since $\psi^*(x) \neq 0$, we can define a holomorphic branch of the nth root in a neighborhood of $\psi^*(x)^n$, such that

$$\psi^*(x) = (\varphi \circ f^*(x))^{1/n}.$$

Then this is a composition of holomorphic functions and is holomorphic.

If we similarly extend our map f^* to a map f by defining $f(x_j) = y$, then f is holomorphic at x_j: chasing arrows through the above commutative diagram and accounting for the value of f at our new point shows that

$$\varphi \circ f \circ \psi^{-1}(z) = z^n$$

By introducing a point x_j for each V_j, and performing this process over each point $y \in R$, we obtain a new Riemann surface X and a holomorphic map $f : X \to Y$. We next show that X is compact. Let Y^- be obtained by removing from Y neighborhoods of the points in R; by choosing these neighborhoods sufficiently small (within coordinate disks), we can assume that the closures of their inverse images in X are compact. Then Y^- is a closed subset of a compact space, and is compact. For each point in Y^-, we choose a neighborhood whose closure is compact and contained in an evenly covered neighborhood; by compactness, we can get an open subcover $U_1, \ldots, U_n$. Then the inverse images of these sets cover $f^{-1}(Y^-)$, and consist of finitely many components homeomorphic to the U_i. In particular, their closures are compact. Thus $f^{-1}(Y^-)$, as a finite union of compact sets, is compact; by our definition of Y^-, the same is true of X.

Finally, we show uniqueness. If $f' : X' \to Y$ is another covering with monodromy conjugate to f, removing the points of X' mapping to branch values gives an unramified covering $f'^* : X'^* \to Y^*$. Theorem 2.12 then shows that there is an isomorphism of coverings $\theta : X^* \to X'^*$. Since $f'^* \circ \theta = f^*$ and f'^* is a local homeomorphism, we can apply the inverse function theorem in a neighborhood of any $x' \in X'^*$ to conclude that $\theta = (f'^*)^{-1} \circ f^*$ for an appropriately locally defined holomorphic inverse of f'^*; thus θ is also holomorphic.

Proposition 1.81 of [3] then states that, if two Riemann surfaces are isomorphic when finitely many points are removed from each, the isomorphism must extend to an isomorphism of the original surfaces. This gives uniqueness.

□

Monodromy is at the heart of this theory, because it allows us to describe topological information in algebraic and combinatorial terms. The above theorem is the most important theorem for us in this regard, because it allows us to define Riemann surfaces and ramified coverings in terms of this algebraic and combinatorial information.

2.4 Algebraicity

Up to this point, we have been talking about Riemann surfaces from the standpoint of complex analysis and topology. However, a remarkable aspect of Riemann surfaces which distinguishes them from other topological objects is that their theory can be recast entirely in the language of algebraic geometry. An introduction to the algebraic side of this theory is given in [11], while a thorough explanation of the technical details can be found in [5].

Definition 2.14. *The n-**dimensional complex projective space** $\mathbb{P}^n$ is the quotient of the space $\mathbb{C}^{n+1}\backslash\{0\}$ by the relation*

$$(z_0, \ldots, z_n) \sim (\lambda z_0, \ldots, \lambda z_n), \quad \lambda \in \mathbb{C}\backslash\{0\}$$

We denote its points by $[z_0 : \cdots : z_n]$.

This is a compact topological space, which follows from the fact that the projection $\mathbb{C}^{n+1}\backslash\{0\} \to \mathbb{P}^n$, when restricted to the (compact) closed unit sphere in $\mathbb{C}^{n+1}$, is still onto. We'd like to talk about Riemann surfaces given by zero loci of polynomials in this new, compact setting, but the value of a polynomial is not well-defined on the equivalence classes of projective space. We turn our focus to **homogeneous polynomials**: those for which every monomial term has the same degree.

If $f(z_0, \ldots, z_n)$ is a homogeneous polynomial of degree d, then

$$f(\lambda z_0, \ldots, \lambda z_n) = \lambda^d f(z_0, \ldots, z_n).$$

So it is well-defined whether f is 0 at a point of $\mathbb{P}^n$.

Definition 2.15. *An **algebraic variety** is a subset of $\mathbb{P}^n$ which is the zero locus of a collection of homogeneous polynomials.*

Note that the term "variety" is often reserved for irreducible varieties: those which cannot be represented as a nontrivial finite union of varieties. For our purposes this makes little difference.

Just as with $\mathbb{P}^1$, $\mathbb{P}^n$ admits a covering by open sets $U_0, \ldots, U_n$, where U_i consists of the points with ith coordinate nonzero, and each one is homeomorphic to $\mathbb{C}^n$, under this map:

$$[z_0 : \ldots : z_n] \mapsto \left(\frac{z_0}{z_i}, \ldots, \frac{z_{i-1}}{z_i}, \frac{z_{i+1}}{z_i}, \ldots, \frac{z_n}{z_i}\right)$$

Under certain conditions, a projective algebraic variety C can be given the structure of a Riemann surface. First, we note that the sets $C \cap U_i$ break C down into several affine curves. The points with ith coordinate nonzero can be put into a canonical form as $[z_0 : \ldots : z_{i-1} : 1 : z_{i+1} : \ldots : z_n]$. Then if C is the zero locus of homogeneous polynomials $f_j(z_0, \ldots, z_n)$, $C \cap U_i$ can be viewed as the affine curve given by the zero locus of the (not necessarly homogeneous) polynomials $f_j(z_0, \ldots, z_{i-1}, 1, z_{i+1}, \ldots, z_n)$.

If the matrix of partial derivatives of the dehomogenized polynomials $f_j(z_0, \ldots, z_{i-1}, 1, z_{i+1}, \ldots, z_n)$ has rank $n-1$ at each point of $C \cap U_i$, then by the Implicit Function Theorem, there is a coordinate at each point such that we can represent all of the other coordinates as holomorphic functions of that coordinate in a neighborhood of our point. In the same way as in the example of affine plane curves at the beginning, we can use these parametrizations to define charts on our curve, and their compatibility follows from the fact that the other coordinate functions are analytic. Similarly, it is straightforward to

check the compatibility of the different charts stemming from $C \cap U_i, C \cap U_j$ for different i and j.

In this case, the variety is called a **smooth algebraic curve**. This highlights an unfortunate conflict of terminology: Riemann surfaces are 2-dimensional real objects (thus the designation of "surface") but also 1-dimensional complex objects (whence "curves").

The starting point for the algebraic geometry of compact Riemann surfaces is that they all arise in this way. This follows from two facts from the more general theory of complex manifolds. (As we have not precisely defined the necessary terminology, we label them as "facts" rather than "theorems".)

Fact ([5, pg. 215])**.** *Every Riemann surface can be holomorphically embedded into $\mathbb{P}^n$ for some n.*

Fact (Chow's Theorem [5, pg. 167])**.** *An analytic subvariety (in particular, submanifold) of $\mathbb{P}^n$ is actually an algebraic variety.*

Together, these facts give us a theorem:

Theorem 2.16. *Every compact Riemann surface is isomorphic to some algebraic variety in $\mathbb{P}^n$ for some n.*

2.4.1 Example: Tori and Elliptic Curves

To prove the above characterization of compact Riemann surfaces is far beyond the scope of this paper, but the reader may feel somewhat cheated by such a result being pulled out of nowhere. A rough example [3, sect. 2.2.1] of how a topologically defined Riemann surface can be given the structure of a variety should hopefully make algebraicity more plausible.

Given two $\mathbb{R}$-linearly independent complex numbers ω_1, ω_2, we can consider the lattice $\Lambda = \{m\omega_1 + n\omega_2 : m, n \in \mathbb{Z}\}$. Then we construct the quotient space of $\mathbb{C}$ obtained from identifying any two points which differ by a point in Λ. This space can also be viewed as a parallelogram with its corners at $0, \omega_1, \omega_2, \omega_1 + \omega_2$ and its opposite sides identified, and we can imagine gluing together the opposite sides to obtain the traditional representation of the torus, as shown in Figure 4.

The torus inherits a Riemann surface structure from the complex plane in a natural way. In particular, the meromorphic functions on the torus correspond to meromorphic functions on $\mathbb{C}$ which are well-defined on the equivalence classes—or equivalently, those which are doubly periodic with periods ω_1 and ω_2.

A straightforward way to construct such functions is by summing a given function over all of the points in $\mathbb{C}$ representing some point of the torus. A slight modification of this approach gives the **Weierstrass $\wp$ function**

$$\wp(z) = \frac{1}{z^2} + \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} \left(\frac{1}{(z - m\omega_1 - n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right)$$

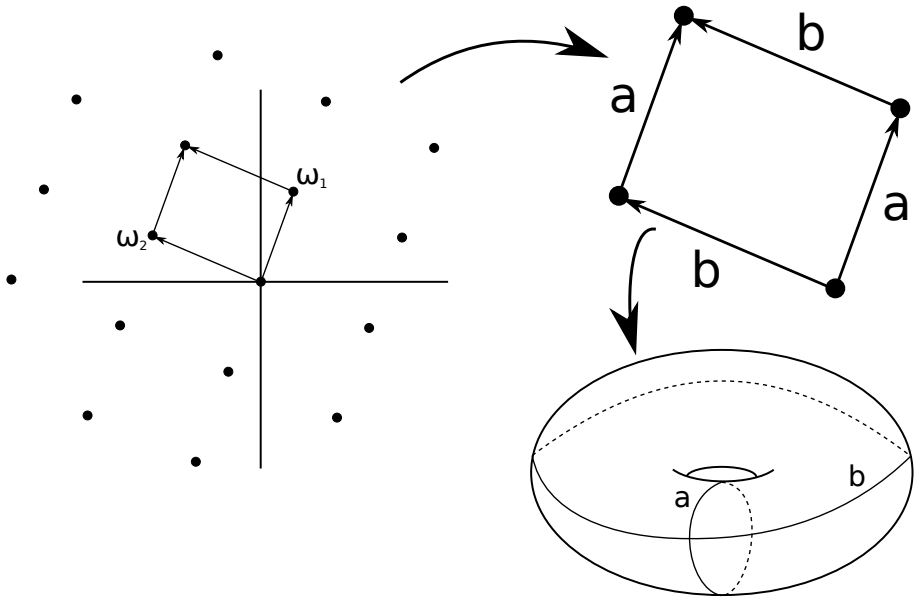


Figure 4: Constructing a torus as a quotient of the complex plane by the lattice generated by ω_1 and ω_2.

and its derivative

$$\wp'(z) = -2 \sum_{m,n \in \mathbb{Z}} \frac{1}{(z - m\omega_1 - n\omega_2)^3}$$

Both functions have poles at 0, of orders 2 and 3 respectively. However, by strategically combining powers of the functions, we can cancel these poles out; examining the Laurent series of $\wp$ and $\wp'$, one can show there exists some constant a such that

$$\wp'^2(z) - 4\wp^3(z) + a\wp(z)$$

is holomorphic everywhere, including the points of Λ.

Now a doubly periodic entire function must be bounded by its maximum value on the base parallelogram; thus by Liouville's theorem, it is constant. So there exists another constant b such that

$$\wp'^2(z) - 4\wp^3(z) + a\wp(z) - b = 0$$

Since $\wp$ and $\wp'$ satisfy the identity, they define a map from our torus into the variety in $\mathbb{P}^3$ defined by $y^2z - 4x^3 - axz^2 - bz^3 = 0$, which turns out to be an isomorphism:

$$z \mapsto [\wp(z) : \wp'(z) : 1]$$

(This definition must be tweaked to account for the pole of $\wp$ and $\wp'$. In a neighborhood of the pole, we express the map as

$$z \mapsto \left[\frac{\wp(z)}{\wp'(z)} : 1 : \frac{1}{\wp'(z)}\right]$$

which is equivalent to the original definition in the rest of the neighborhood.)

Thus we see why tori are equivalent to algebraic curves: by constructing meromorphic functions on them, we can then establish a purely algebraic relationship between these functions, such that they give a morphism from the torus to the algebraic curve defined by that relationship. The problem of constructing "enough" meromorphic functions is central to proving algebraicity.

Algebraic curves of genus 1, of the sort just constructed, are called **elliptic curves**, despite not being ellipses.

2.4.2 Morphisms of Varieties

Once we know every Riemann surface is isomorphic to an algebraic curve, we can examine what holomorphic maps and meromorphic functions become in this new context.

In analogy with $\mathbb{P}^1$, we start out by considering meromorphic functions given by rational functions. We saw above that polynomials cannot be evaluated at points of projective space in a well-defined way. Even if the polynomial p is homogeneous of degree d, scaling the input by λ still scales the output by λ^d.

However, if we attempt to evaluate a ratio of two homogeneous polynomials of the same degree at a point of $\mathbb{P}^n$, these scale factors will cancel out, and the result will be well-defined. Thus when we talk about rational functions on $\mathbb{P}^n$ or a curve embedded in it, we refer to quotients of two homogeneous polynomials of the same degree.

It is straightforward to verify [11, Proposition II.2.11–12] that restricting such a rational function to a projective algebraic curve gives a meromorphic function, as long as the denominator is not identically zero on the curve. The second central result on algebraicity of meromorphic functions is:

Theorem 2.17 ([5, pg. 168])**.** *Every meromorphic function on a projective algebraic curve is the restriction of a rational function.*

We can work from this to similarly characterize maps between projective curves in terms of polynomials. Suppose $\varphi : S \to T$ is a holomorphic map between $S \subset \mathbb{P}^n$ and $T \subset \mathbb{P}^m$. Then we can write

$$\varphi(p) = [\varphi_0(p) : \ldots : \varphi_m(p)].$$

where the φ_i are defined only up to multiplication of all of them by a constant. Then the functions φ_i/φ_j are well-defined meromorphic functions on S, since they come from composing φ with meromorphic functions on T. Thus they can be given by rational functions.

Now we fix an index j and restrict our attention to the subset $\varphi^{-1}(U_j) = \{p \in S : \varphi_0(p) \neq 0\}$. On this set, the meromorphic functions φ_i/φ_j have no poles, and so they can be represented by ratios of polynomials f_i/g_i such that g_i is never 0 on this set:

$$\varphi(p) = \left[\frac{f_0(p)}{g_0(p)} : \ldots : \frac{f_{j-1}(p)}{g_{j-1}(p)} : 1 : \frac{f_{j+1}(p)}{g_{j+1}(p)} : \ldots : \frac{f_m(p)}{g_m(p)}\right]$$

By multiplying through by the polynomial $g_0 g_1 \ldots g_{j-1} g_{j+1} \ldots g_m$ to clear denominators, we can obtain a representation of φ consisting entirely of polynomials in the coordinates of p. (Here we use the fact that the g_i are nonzero on $\varphi^{-1}(U_j)$.)

We have shown that every morphism of projective curves is given locally (on a finite collection of open sets) by tuples of homogeneous polynomials of the same degree. Our preceding characterization of meromorphic functions is a special case of this, under the identification with morphisms to $\mathbb{P}^1$. Furthermore, each of the sets $\varphi^{-1}(U_j)$ is dense in S, since $\varphi^{-1}(U_j)$ is defined by the nonvanishing of a meromorphic function, whose zeros are isolated. Thus a morphism of Riemann surfaces is determined by its polynomial representation on just one of these sets, once one is known.

There are various conditions that must be checked to ensure that any such collection of tuples actually gives a morphism between two specified curves, which can be phrased in terms of equations between the relevant polynomials or ideals. Rather than working this out in full detail (for which we refer the reader to Corollaries 2.7–2.9 of [4]), we will state it as part of a General Principle, followed by a couple of examples.

General Principle. *For a collection of tuples of homogeneous polynomials of the same degree to define a map between algebraic curves S and T can be characterized by polynomial relations between the polynomials defining the map, S, and T. The same is true of determining whether such a collection defines a map of ramified coverings of a given surface, and determining whether a map is an isomorphism.*

Example. Suppose $S \subset \mathbb{P}^n$ is defined by the vanishing of polynomials P_i and $T \subset \mathbb{P}^m$ is defined by the vanishing of polynomials Q_j. Consider homogeneous polynomials $F_0, \ldots, F_m$ of the same degree. If U is the open subset of S where these are not simultaneously zero, we can consider the map $U \to \mathbb{P}^m$ given by $p \mapsto [F_0(p) : \ldots : F_m(p)]$. For this to actually be a map into T, the polynomials $Q_j(F_0, \ldots, F_m)$ must identically vanish on U.

Again, U is dense in S (any point at which the P_i are identically zero is arbitrarily close to ones for which they aren't) and so equivalently these polynomials must vanish on all of S. One form of the Nullstellensatz [1, Exercise 7.14] states that a polynomial will identically vanish on S exactly when some power of that polynomial lies in the ideal generated by the P_i. Thus the criterion that our F_k define a map into T comes down to the existence of an integer e and polynomials C_i such that

$$(Q_j(F_0, \ldots, F_m))^e = \sum C_i P_i$$

This is the kind of polynomial relation we are talking about above. A similar such relation can be constructed to determine when tuples of polynomials on several different open subsets are compatible on their overlaps.

With this in place, the category of compact Riemann surfaces metamorphoses into something entirely algebraic. The strength of having introduced Riemann surfaces the way we did is that topological concepts such as genus and monodromy are still available to us.

3 Belyi's Theorem

As mentioned above, every Riemann surface can be realized as a projective algebraic curve, which is already kind of miraculous. This fact allows us to translate between the realms of complex analysis and algebra in unexpected ways, one of which is the topic of this section and the starting point for the general theory of dessins d'enfants.

We begin with an algebraic definition. Given a subfield $K \subset \mathbb{C}$, we say that a Riemann surface is **defined over** K if it is isomorphic to a curve given by equations with coefficients in K. In keeping with our focus on ramified coverings rather than individual surfaces, we can similarly say that a covering $f : X \to \mathbb{P}^1$ is defined over K if it is isomorphic to a covering $f' : X' \to \mathbb{P}^1$ where X' is a curve with coefficients in K and f' is given by polynomials with coefficients in K.

The condition that this notion is only defined up to isomorphism is necessary for it to be well-defined on the category of Riemann surfaces, but it is somewhat loose, and can lead to counterintuitive results. For example, the curves $y^2z - x^3 - az^3 = 0$ are isomorphic for all $a \neq 0$, with an isomorphism to the curve $y^2z - x^3 - z^3 = 0$ given by

$$[x : y : z] \mapsto \left[x : \frac{y}{\sqrt[6]{a}} : \sqrt[3]{a}z\right]$$

Thus, even though a could be anything, all of these curves are defined over $\mathbb{Q}$.

From the point of view of number theory, a question that naturally arises is when a curve is defined over a **number field**: a finite extension of $\mathbb{Q}$. It will suffice to ask when a curve is defined over $\overline{\mathbb{Q}}$: the algebraic closure of $\mathbb{Q}$, or the field of **algebraic numbers**.

Thus we introduce an algebraic aspect to Riemann surfaces which would seem to be unrelated to the analytic and topological theory reviewed above. The remarkable conclusion of Belyi's theorem is that it isn't.

Theorem 3.1 (Belyi's Theorem)**.** *Let S be a Riemann surface. The following are equivalent:*

1. *S is defined over $\overline{\mathbb{Q}}$.*
2. *There is a holomorphic map $f : S \to \mathbb{P}^1$ which ramifies over at most 3 points.*

In this case the covering f is also defined over $\overline{\mathbb{Q}}$.

By composing with an appropriate Möbius transformation, we can send the 3 branch values to any three points. The standard choice is $\{0, 1, \infty\}$, and we will assume that such a map ramifies over these points unless stated otherwise. We will also refer to maps which specifically ramify over $\{0, 1, \infty\}$, when using the language of meromorphic functions, as **Belyi functions**.

This section is devoted to a proof of this theorem. We follow the treatment of [3] and [4], in particular their proof of $(1) \Rightarrow (2)$, which avoids some technical background in algebraic geometry used by more standard proofs.

3.1 Galois Actions

Central to identifying what is special about Riemann surfaces defined over $\overline{\mathbb{Q}}$ is understanding how the field $\overline{\mathbb{Q}}$ sits inside $\mathbb{C}$. To this end, we will use tools from the theory of fields, specifically $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$, the group of field automorphisms of $\mathbb{C}$. The first step is to construct natural actions of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ on the various objects we have been considering—polynomials, points of $\mathbb{P}^n$, surfaces, coverings. The guiding principle is that we should be able to treat these objects, with regard to the Galois action, as we treat numbers.

(With the exception of one lemma, the content of this section is not needed for the proof of the (1) $\Longrightarrow$ (2) direction of Belyi's theorem. The reader who wants to get a feel for the proof may read the following section first.)

Suppose $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ is any field automorphism. First, for any point $p = [p_0 : \ldots : p_n] \in \mathbb{P}^n$, we define $p^\sigma = [p_0^\sigma : \ldots : p_n^\sigma]$. This action is well-defined, since if we scale all coordinates of p by some λ, the coordinates of p^σ are scaled by λ^σ.

From this, we also get a Galois action on subsets of $\mathbb{P}^n$. We can apply this to Riemann surfaces in $\mathbb{P}^n$, but since the Galois action is something purely algebraic, there is no a priori reason why the result should have the topological structure of a Riemann surface. The key step is to use the interpretation of Riemann surfaces as algebraic curves.

So, given a polynomial $f(z_0, \ldots, z_n) = \sum a_{i_0 \cdots i_n} z_0^{i_0} \cdots z_n^{i_n}$, we define $f^\sigma = \sum \sigma(a_{i_0 \cdots i_n}) z_0^{i_0} \cdots z_n^{i_n}$. Then if S is a Riemann surface defined by the vanishing of homogeneous polynomials $f_1, \ldots, f_k$, it follows from the fact that σ is an automorphism that S^σ is exactly the zero locus of the polynomials $f_1^\sigma, \ldots, f_k^\sigma$. The Galois action on polynomials commutes with formal differentiation, and so the conditions on rank and nonvanishing of derivatives that imply the zero locus of $f_1, \ldots, f_k$ is a Riemann surface also imply that S^σ is.

Similarly, we can define an action on morphisms of Riemann surfaces embedded in projective space: given $S \subset \mathbb{P}^n$, $S' \subset \mathbb{P}^m$, and $f : S \to S'$, we define $f^\sigma : S^\sigma \to S'^\sigma$ by $f^\sigma = \sigma \circ f \circ \sigma^{-1}$. Again, this can be defined in terms of polynomials. Using the characterization of section 2.4.2, suppose f is given on the sets U_i by tuples of polynomials $[F_{i,0} : \ldots : F_{i,m}]$, such that U_i is the (open) set on which the $F_{i,j}$ are not simultaneously 0. Then U_i^σ is the set on which $F_{i,j}^\sigma$

are not simultaneously 0, and f^σ is given by $F_{i,j}^\sigma$ on this set; from this, we see that f^σ is also a morphism of Riemann surfaces.

The above definitions make the action of any element of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ into a functor from the category of Riemann surfaces embedded in projective space to itself, since $(f \circ g)^\sigma = f^\sigma \circ g^\sigma$. This has a couple of immediate but important consequences:

- If $f : S \to S'$ is an isomorphism of Riemann surfaces, then $f^\sigma : S^\sigma \to S'^\sigma$ is also an isomorphism (with inverse $(f^{-1})^\sigma$), and so our action is well-defined on isomorphism classes. In particular, we can talk about applying it to any compact Riemann surface, without worrying about the embedding into projective space.
- If $f : S \to T$ and $f' : S' \to T$ are ramified coverings of a Riemann surface T, and $\varphi : S \to S'$ is a morphism of coverings (that is, such that $f' \circ \varphi = f$) then φ^σ is a morphism between the coverings $f^\sigma : S^\sigma \to T^\sigma$ and $f'^\sigma : S'^\sigma \to T^\sigma$.

Finally, we note one more property of the Galois action which is critical to the coming proofs.

Proposition 3.2. *For any map $f : S \to S'$ of Riemann surfaces and $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$, $\deg(f) = \deg(f^\sigma)$. Additionally, f ramifies over $a \in S'$ if and only if f^σ ramifies over a^σ.*

Proof. Given $p \in S$ and $a \in S'$, we have $f(p) = a$ if and only if $f^\sigma(p^\sigma) = a^\sigma$, where the "if" direction follows from applying σ^{-1}. Thus σ gives a bijection between the fibers $f^{-1}(\{p\})$ and $(f^\sigma)^{-1}(\{p^\sigma\})$. All but finitely many fibers of f (resp. f^σ) will have size $\deg(f)$ (resp. $\deg(f^\sigma)$), thus the two degrees must be equal. Then f ramifies over a if and only if $|f^{-1}(\{a\})| < \deg(f)$ if and only if $|(f^\sigma)^{-1}(\{a^\sigma\})| < \deg(f^\sigma)$ if and only if f^σ ramifies over a^σ. □

3.2 (1) $\Longrightarrow$ (2)

The direction of Belyi's theorem which Belyi was actually responsible for proceeds in two steps. The idea is to begin with a meromorphic function having an arbitrary number of algebraic branch values, and then modify it by composing with other functions:

- first, to make all of the branch values rational or ∞;
- second, to collapse these rational branch values together until there are only 3 of them, at $0, 1, \infty$.

In each step, the proof amounts to very little beyond a clever choice of functions to compose with.

Central to this proof, then, is the way branch values of a function transform under composition with other functions.

Lemma 3.3. *If $f : S_1 \to S_2$ and $g : S_2 \to S_3$ are two morphisms of Riemann surfaces, having branch values $\{w_1^f, \ldots, w_m^f\}$ and $\{w_1^g, \ldots, w_n^g\}$ respectively, then the branch values of $g \circ f$ are*

$$\{w_1^g, \ldots, w_n^g\} \cup \{g(w_1^f), \ldots, g(w_m^f)\}$$

Proof. This can be seen from Theorem 2.5, which characterizes a branch value as a point, the fiber over which is smaller than usual (that is, smaller than the degree of the morphism). Since $(g \circ f)^{-1}(\{p\}) = f^{-1}(g^{-1}(\{p\}))$, p will be a branch value if and only if either $g^{-1}(\{p\})$ is smaller than $\deg g$ (so it is a branch value of g) or for some $q \in g^{-1}(p)$, $f^{-1}(q)$ is smaller than $\deg f$ (so p is $g(q)$ for q a branch value of f). This is the characterization we wanted. □

We need to choose some nonconstant meromorphic function on S to start with, but all we need is for it to have algebraic branch values.

Lemma 3.4. *Let S be a Riemann surface defined over an algebraically closed field $K \subset \mathbb{C}$, and let $f : S \to \mathbb{C} \cup \{\infty\}$ be a meromorphic function, given on a dense set by a rational function with coefficients in K. Then the branch values of S lie in $K \cup \{\infty\}$.*

Proof. (from [12]) First, note that there are finitely many branch values. The definition of ramification means that the points of ramification of f are given locally by the vanishing of an analytic function (the derivative of the coordinate representation of f.) Thus they are isolated. Then we can use the compactness of S: we can form an open cover of S by the complement of the set of ramification points and neighborhoods around each ramification point small enough to contain no others, and this can only have a finite subcover if there are finitely many ramification points, and thus finitely many branch values.

Now consider the group $\mathrm{Gal}(\mathbb{C}/K)$ of automorphisms of $\mathbb{C}$ fixing K. Let σ be an arbitrary automorphism from this group. Then if f has branch values $\{a_1, \ldots, a_n\}$, by Proposition 3.2, $f^\sigma : S^\sigma \to \mathbb{C} \cup \{\infty\}$ has branch values $\{a_1^\sigma, \ldots, a_n^\sigma\}$.

On the other hand, since S is defined by polynomials with coefficients in K, $S^\sigma = S$. Similarly, since f is defined by a rational function with coefficients in K, $\sigma \circ f \circ \sigma^{-1} = f$. Thus the action of $\mathrm{Gal}(\mathbb{C}/K)$ permutes the branch values $\{a_1, \ldots, a_n\}$.

Now, if $a_i \notin K$, then since K is algebraically closed, a_i must be transcendental over K. On the other hand, the orbit of a transcendental number under the action of $\mathrm{Gal}(\mathbb{C}/K)$ must be infinite, as there exist infinitely many transcendental elements over K and elements $\mathrm{Gal}(\mathbb{C}/K)$ sending a_i to any one of them [2, pg. V.2, Proposition 9]. This is a contradiction. □

So we only need a nonconstant rational function with coefficients in $\overline{\mathbb{Q}}$. This is easy enough to find: for example, choosing some coordinate (WLOG z_n) which is not identically zero on S, one of the rational functions $[z_0 : \ldots z_n] \mapsto z_i/z_n$ (WLOG z_0/z_n) must be nonconstant. We let $c_0 := z_0/z_n$. Then this will ramify over algebraic numbers $\{\mu_1, \ldots, \mu_n\}$, as well as possibly ∞.

3.2.1 Step 1

If all the μ_k are rational, we can move on to the second step. Otherwise, let f_1 be the minimal polynomial of all the irrational μ_k. Then it sends all of the irrational branch values of c_0 to 0, or ∞; thus by Lemma 3.3, the branch values of $f_1 \circ c_0$ are just those of f_1, along with 0, ∞, and the values $f_1(\mu_k)$ for μ_k rational, which will themselves be rational. We recall that f_1 ramifies at ∞ and the roots of f_1', which we denote by $\{\beta_1, \ldots, \beta_m\}$; thus the finite branch values of f_1 are the values $f_1(\beta_k)$.

Now, these branch values may themselves be irrational. (If they are all rational, we can proceed to the next step.) But we claim that we are in a better position than we started out in. Specifically, if we do the same construction again and let f_2 be the minimal polynomial of the irrational branch values of $f_1 \circ c_0$, then we claim that $\deg f_2 < \deg f_1$.

If we let h be the minimal polynomial of the irrational β_k, then since $f'(\beta_k) = 0$ for all k, certainly $\deg h \leq \deg f_1' < \deg f_1$. The roots of h are all simple roots, given by the Galois conjugates of the β_k, those numbers of the form $\sigma(\beta_k)$ for $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Similarly, the (simple) roots of f_2 consist of the distinct values of $\sigma(f_1(\beta_k)) = f_1(\sigma(\beta_k))$; but then there can only be so many such values as there are distinct values of $\sigma(\beta_k)$, implying $\deg f_2 \leq \deg h < \deg f_1$.

Knowing this, we can repeat the process described above. We consider the map $f_2 \circ f_1 \circ c_0$; if it has irrational branch values, then they are branch values of f_2, and, letting f_3 be their minimal polynomial, we have $\deg f_3 < \deg f_2$. We then shift our focus to $f_3 \circ f_2 \circ f_1 \circ c_0$, and so on. Since the degree of the minimal polynomial strictly decreases with each iteration, the process must eventually terminate in a meromorphic function $g = f_k \circ f_{k-1} \circ \cdots \circ f_1 \circ c_0$, all of whose finite branch values are rational.

3.2.2 Step 2

Once we have our function g, we can assume without loss of generality that its branch values include 0, 1, and ∞, by composing with a Möbius transformation (with rational coefficients). In fact, we can do slightly better, and assume that one of the other branch values lies between 0 and 1. Given any rational number $\neq 0, 1$, we can apply some combination of the Möbius transformations $1-z$ and $1/z$ to shift it into this interval, while merely permuting the values $\{0, 1, \infty\}$.

Thus we assume that our function g has branch values $\{0, 1, \infty, \lambda_1, \ldots, \lambda_n\}$, where the λ_i are rational and $0 < \lambda_1 < 1$. We can specifically write $\lambda_1 = m/(m+k)$ for positive integers m and k. Then we deploy the following magical polynomial:

$$P_{m,k}(z) = \frac{(m+k)^{m+k}}{m^m k^k} z^m (1-z)^k$$

Note that

$$P'_{m,k}(z) = \frac{(m+k)^{m+k}}{m^m k^k}(m(1-z)+kz)z^{m-1}(1-z)^{k-1}$$

and so its roots, the ramification points of $P_{m,k}$, are 0, 1 and $m/(m+k) = \lambda_1$. The corresponding branch values are $P_{m,k}(0) = P_{m,k}(1) = 0, P_{m,k}(\lambda_1) = 1$, and ∞.

We now apply Lemma 3.3 to $P_{m,k} \circ g$. Applying $P_{m,k}$ to the branch values of g collapses 0 and 1 to 0, sends λ_1 to 1 and ∞ to ∞. Then since the branch values of $P_{m,k}$ are $\{0, 1, \infty\}$, they do not contribute anything new to the set, and the set of branch values of $P_{m,k} \circ g$ is one smaller than that of g. We can then repeat this process until $0, 1, \infty$ are the only branch values, at which point we are done.

3.2.3 A Note on Computation and an Example

In principle, this formula gives us an algorithm for finding a Belyi function on an algebraic curve. However, because the degrees of the polynomials $P_{m,k}$ are dependent on the numerator and denominator of the branch values, the polynomials themselves are scaled by factors with very large numerator and denominator, and dealing with multiple branch values requires composing several such polynomials together, the degree of the resulting polynomial will typically be comically huge. For example, Example 2.6.4 in [8] details how, to turn a function with branch values $0, 1/5, 2/5, 3/5, 4/5, 1, \infty$ into a Belyi function in this manner, one must compose with a polynomial of degree approximately $10^{5\times 10^{25}}$. Belyi later gave an alternative argument which produces more reasonable polynomials [3, pg. 175].

Nonetheless, we can see how this algorithm works in a simple example.

Example. Consider the elliptic curve given by $y^2z = x(x-z)(x-\sqrt{2}z)$ in $\mathbb{P}^2$, which is clearly defined over $\overline{\mathbb{Q}}$. We start by examining the ramification of c_0.

It is helpful to break down the projective description of this curve: there is one "point at ∞" having $z = 0$, at $[0:1:0]$, and all other points can be written in the form $[x:y:1]$, with $y^2 = x(x-1)(x-\sqrt{2})$. We see that for most finite values of x, there are two points $[x:y:1]$ in the curve, with y given by either of the square roots of $x(x-1)(x-\sqrt{2})$. The finite branch values of c_0 are then those x for which there is only one such point, which are exactly the values $0, 1, \sqrt{2}$, for which $x(x-1)(x-\sqrt{2})$ has only one square root. Additionally, the fiber $c_0^{-1}(\{\infty\})$ can only contain $[0:1:0]$, so ∞ is also a branch value.

In step 1, the only irrational branch value is $\sqrt{2}$, which has minimal polynomial $w^2 - 2$. This has -2 and ∞ as its only branch values, and it sends the existing branch values $0, 1, \sqrt{2}, \infty$ to $-2, -1, 0, \infty$. Thus the composition $c_0^2 - 2$ ramifies over -2, -1, 0, and ∞.

Now these are all rational, so we can move on to step 2. Applying the Möbius transformation $-1/w$ sends the values to $1/2$, 1, ∞, and 0. Then since $1/2 = 1/(1+1)$, we can apply the polynomial $P_{1,1}(w) = 4w(1-w)$, which will give us a function ramified over $0, 1, \infty$.

The chain of compositions and resulting function are thus:

$$c_0 \xrightarrow{w^2-2} c_0^2 - 2 \xrightarrow{-1/w} \frac{-1}{c_0^2-2} \xrightarrow{4w(1-w)} \frac{-4(c_0^2-1)}{(c_0^2-2)^2}$$

Now, we move on to the other direction of the theorem, which is entirely different.

3.3 An Alternative Criterion for (1)

The technical details of this direction are somewhat involved, but the proof we present here is based around a simple guiding principle: the idea that we can use the action of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ to determine when a covering is defined over $\overline{\mathbb{Q}}$ in the same way as we can use it to determine when a number is in $\overline{\mathbb{Q}}$.

Basic field theory then gives a clean separation of $\mathbb{C}$ into algebraic numbers and transcendental numbers using this group. If $z \in \mathbb{C}$ is transcendental, then there is an automorphism $\mathbb{C} \to \mathbb{C}$ sending it to any other transcendental number [2, pg. V.2, Proposition 9], and so the transcendental numbers form one uncountable orbit under the action of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$. On the other hand, the orbit of an algebraic number under this action consists of finitely many numbers: its **Galois conjugates**, the roots of its minimal polynomial.

The analogous criterion for coverings is:

Theorem 3.5. *For a ramified covering $f : S \to \mathbb{P}^1$, the following are equivalent:*

1. *f is defined over $\overline{\mathbb{Q}}$.*
2. *The orbit of f under the action of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ consists of finitely many isomorphism classes.*

One can show 1 $\Rightarrow$ 2 by noting that the finitely many polynomials defining S and f have finitely many algebraic coefficients, which will themselves have finite orbits under $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$. But the direction we need is that 2 $\Rightarrow$ 1, which we now prove.

3.4 Proving the Criterion

We assume that the orbit of $f : S \to \mathbb{P}^1$ under $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ has finitely many isomorphism classes.

This condition gives us a profusion of covering isomorphisms $\varphi : S \to S'$, and so the central idea of the proof is to transform one of these isomorphisms such that the other side is defined over $\overline{\mathbb{Q}}$. This will proceed in two steps. The first is to find an isomorphism such that S and S' are independent from each other in a certain algebraic sense. The second is to use this independence

to make small adjustments to the coefficients of the polynomials defining $f' : S' \to \mathbb{P}^1$ and φ while leaving f untouched. While the underlying ideas are simple, there are some algebraic details we must wade through.

3.4.1 Transcendence

Let k be a subfield of $\mathbb{C}$. Recall that $a \in \mathbb{C}$ is called algebraic over k if it is the root of some polynomial with coefficients in k, and transcendental if it isn't. Alternatively, a is algebraic if and only if the map $k[x] \to \mathbb{C}$ defined by $p(x) \mapsto p(a)$ has nontrivial kernel, and transcendental if and only if this map is injective.

More generally, we say that a set $\{a_1, \ldots, a_n\} \subset \mathbb{C}$ is **algebraically independent over** k [7, Definition VI.1.1] if the elements satisfy no polynomial equation with coefficients in k, so that the map $k[x_1, \ldots, x_n] \to \mathbb{C}$ defined by $p \mapsto p(a_1, \ldots, a_n)$ is injective. In the absence of a specified field, k refers to $\mathbb{Q}$ by default.

Algebraically independent sets are plentiful. Given an algebraically independent set $\{a_1, \ldots, a_n\}$, we claim it is always possible to add some a_{n+1} which is transcendental over $\mathbb{Q}(a_1, \ldots, a_n)$, and thus extends it to a larger independent set. If this were not the case, every element of $\mathbb{C}$ would be algebraic over $\mathbb{Q}(a_1, \ldots, a_n)$, so that $\mathbb{C}$ would be the algebraic closure of this field; but $\mathbb{Q}(a_1, \ldots, a_n)$ is countable, so its algebraic closure must be as well, a contradiction.

Algebraically independent sets are also flexible with regards to the action of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$. If $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ are two algebraically independent sets, then there exists an automorphism $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ such that $\sigma(a_i) = b_i$. This follows by extending the isomorphism $\mathbb{Q}(a_1, \ldots, a_n) \cong \mathbb{Q}(b_1, \ldots, b_n)$ to an automorphism of $\mathbb{C}$, as described in [2, pg. V.107, Corollaire 2].

Given polynomials defining S and $f : S \to \mathbb{P}^1$, let $\{c_1, \ldots, c_n\}$ be a maximal algebraically independent subset of their coefficients. Based on the two facts above, we claim:

Lemma 3.6. *There exists some $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ such that $\{c_1, \ldots, c_n, \sigma(c_1), \ldots, \sigma(c_n)\}$ are algebraically independent and f is isomorphic to f^σ.*

We use these facts to set up the morphism of coverings which we will manipulate to get an isomorphism with a covering defined over $\overline{\mathbb{Q}}$. Abusing terminology somewhat, we will say two sets $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ are algebraically independent of each other if they are disjoint and their union is algebraically independent.

Given polynomials defining S and $f : S \to \mathbb{P}^1$, let $\{c_1, \ldots, c_n\}$ be a maximal algebraically independent subset of their coefficients. Our results on the flexibility and quantity of algebraically independent sets then show that we can find an infinite collection of $\sigma_\alpha \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ such that the sets $\{\sigma_\alpha(c_1), \ldots, \sigma_\alpha(c_n)\}$ are pairwise algebraically independent of each other, by extending $\{c_1, \ldots, c_n\}$ to an infinite algebraically independent set and choosing automorphisms sending the c_i to different elements of this set. Our assumed finiteness condi-

tion and the pigeonhole principle then imply that there must exist two such automorphisms σ', σ'' such that $f^{\sigma'}$ and $f^{\sigma''}$ are isomorphic. Then letting $\sigma = \sigma'^{-1} \circ \sigma''$ and applying σ'^{-1} to this isomorphism gives an isomorphism $\varphi : S \to S^{\sigma}$ of the coverings f and f^{σ}.

3.4.2 Specialization

Having the above isomorphism puts us in a powerful position: the independence of (some of) the coefficients on one side from those on the other will allow us to manipulate the two sides separately.

We have previously seen that we can transform isomorphisms by the Galois action. However, this can only take transcedental elements to transcendental elements, which is not what we need. Instead, we will use a more general sort of tweaking called specialization. We have seen that algebraic independence of a set $\{a_1, \ldots, a_n\}$ is characterized by the evaluation map $\mathbb{Q}[x_1, \ldots, x_n] \to \mathbb{Q}[a_1, \ldots, a_n]$ being an isomorphism. Then for *any* elements $q_1, \ldots, q_n \in \mathbb{C}$, we have a **specialization**: a homomorphism $s : \mathbb{Q}[a_1, \ldots, a_n] \to \mathbb{C}$ sending $a_i \mapsto q_i$, corresponding to the analogous map from the polynomial ring. As with the Galois action, for a polynomial $p(x) = \sum a_i x^i$, we define $p^s(x) = \sum s(a_i) x^i$, and this gives a homomorphism $\mathbb{Q}[a_1, \ldots, a_n][x] \to \mathbb{C}[x]$.

There are two hurdles we must clear to use specialization on our isomorphism. The first is that, if there is some algebraic relation between the q_i, s will have a nontrivial kernel. Thus we can't extend it to a map $\mathbb{Q}(a_1, \ldots, a_n) \to \mathbb{C}$, as it might map valid denominators to 0. The best we can do is extend s to a map $\mathbb{Q}[a_1, \ldots, a_n]_{\ker(s)} \to \mathbb{C}$, where $\mathbb{Q}[a_1, \ldots, a_n]_{\ker(s)}$ is the localization at $\ker(s)$, the subring of rational functions whose denominators do not lie in $\ker(s)$ (thus avoiding division by 0). (Though we only require this direct definition, context for localization can be found in Chapter 3 of [1].)

Secondly, if we want to apply a specialization to the coefficients of our coverings, we need to account for the coefficients which lie outside the algebraically independent sets chosen above. Let K be the field generated by *all* the coefficients of the defining polynomials of f; then by the assumption of maximality, the remaining coefficients are algebraic over $\mathbb{Q}(c_1, \ldots, c_n)$. So K is a finitely generated algebraic (thus finite) extension of $\mathbb{Q}(c_1, \ldots, c_n)$, and by the Primitive Element Theorem [7, Theorem V.6.16], $K = \mathbb{Q}(c_1, \ldots, c_n, u)$ for some single u algebraic over $\mathbb{Q}(c_1, \ldots, c_n)$.

The question then becomes: given some element v algebraic over $\mathbb{Q}(a_1, \ldots, a_n)$, can we extend s to the ring $\mathbb{Q}[a_1, \ldots, a_n, v]$? To account for the algebraicity of v, we will consider the minimal polynomial $p_1(x)$ of v over $\mathbb{Q}(a_1, \ldots, a_n)$. However, to keep its coefficients in the ring $\mathbb{Q}[a_1, \ldots, a_n]$, we let d be the least common denominator of the coefficients of $p_1(x)$ (well-defined up to multiplication by a unit, because $\mathbb{Q}[a_1, \ldots, a_n]$ is a unique factorization domain), and let $p(x) := dp_1(x)$. This has relatively prime coefficients in $\mathbb{Q}[a_1, \ldots, a_n]$.

Proposition 3.7. *Suppose $s_0 : \mathbb{Q}[a_1, \ldots, a_n] \to \mathbb{C}$ is a specialization. Then it has an extension $s : \mathbb{Q}[a_1, \ldots, a_n, v] \to \mathbb{C}$ with $s(v) = v_1$ if and only if v_1 is a root of $p^{s_0}(x)$.*

Proof. If we let x be an indeterminate and $\pi : \mathbb{Q}[a_1, \ldots, a_n, x] \to \mathbb{Q}[a_1, \ldots, a_n, v]$ be the map setting $x = v$, then maps $s : \mathbb{Q}[a_1, \ldots, a_n, v] \to \mathbb{C}$ correspond precisely to maps $\widetilde{s} : \mathbb{Q}[a_1, \ldots, a_n, x] \to \mathbb{C}$ which vanish on $\ker(\pi)$, via the following commutative diagram:

$$\begin{array}{ccc} \mathbb{Q}[a_1, \ldots, a_n, x] & \xrightarrow{\widetilde{s}} & \mathbb{C} \\ \downarrow{\scriptstyle \pi} & \nearrow_{s} & \\ \mathbb{Q}[a_1, \ldots, a_n, v] & & \end{array}$$

This follows from the first isomorphism theorem and the universal property of quotient rings, applied to $\mathbb{Q}[a_1, \ldots, a_n, v] \cong \mathbb{Q}[a_1, \ldots, a_n, x]/\ker(\pi)$.

In this case, the kernel of the evaluation map $\mathbb{Q}(a_1, \ldots, a_n)[x] \to \mathbb{Q}(a_1, \ldots, a_n, v)$ is $(p_1(x))$. Then it follows that $\ker(\pi) = (p_1(x)) \cap \mathbb{Q}[a_1, \ldots, a_n, x]$, and we claim this ideal is actually $(p(x))$. Indeed, suppose $r(x) = p_1(x)q(x) = p(x)(q(x)/d) \in \mathbb{Q}[a_1, \ldots, a_n, x]$. Then Gauss's lemma [9, Corollary IV.2.2] implies that, since $p(x)$ has relatively prime coefficients, $q(x)/d \in \mathbb{Q}[a_1, \ldots, a_n, x]$ as well, and so $r(x) \in (p(x))$.

Thus we want to consider maps $\widetilde{s} : \mathbb{Q}[a_1 \ldots, a_n, x] \to \mathbb{C}$ which restrict to s_0 on $\mathbb{Q}[a_1, \ldots, a_n]$ and which are 0 on $p(x)$. That is, $\widetilde{s}(p(x)) = p^{s_0}(\widetilde{s}(x)) = 0$, so the possible values for $\widetilde{s}(x)$, and thus $s(v)$, are the roots of p^{s_0}. □

At this point, we can lay out the proof strategy in more detail. We return to the notation of the previous subsection, where $\{c_1, \ldots, c_n\}$ is a maximal algebraically independent subset of the coefficients of S and f, $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ is such that $\{c_1, \ldots, c_n, \sigma(c_1), \ldots, \sigma(c_n)\}$ are algebraically independent, and $\varphi : S \to S^\sigma$ is an isomorphism.

By the General Principle of section 2.4.2, the facts that f and f^σ are coverings and that φ is an isomorphism between them are encoded by algebraic relations between their defining polynomials and various other auxiliary polynomials. We refer to the collection of these polynomials as $\mathcal{P}$. We take a subset $\{c_{n+1}, \ldots, c_m\}$ of the coefficients of all these polynomials, maximal with respect to the property that

$$A = \{c_1, \ldots, c_n, \sigma(c_1), \ldots, \sigma(c_n), c_{n+1}, \ldots, c_m\}$$

is an algebraically independent set. By the same reasoning used above, the field L generated by *all* of the coefficients of the polynomials in $\mathcal{P}$ is an algebraic extension $\mathbb{Q}(A, u') \supset \mathbb{Q}(A)$.

Our plan is then to choose a specialization $s : \mathbb{Q}[A, u'] \to \mathbb{C}$ such that $s(c_i) = c_i$ for $i \leq n$, while the numbers $q_i = s(\sigma(a_i))$ are algebraic. Then we'd like to apply s to the polynomials in $\mathcal{P}$. Crucially, since s acts homomorphically, this would preserve all of the relations between polynomials involved in the General Principle, and so we would get an isomorphism φ^s between f^s and $(f^\sigma)^s$, where $f^s = f$ and $(f^\sigma)^s$ is defined over $\overline{\mathbb{Q}}$.

However, the problem with this plan as stated is that the coefficients of the polynomials in $\mathcal{P}$ lie in $\mathbb{Q}(A, u')$, and as stated above there is generally no way of extending a specialization to this whole field. Additionally, in order to ensure that s actually fixes f, we must know that it fixes the remaining algebraic generator u of K as well as the c_i. To fix these issues, we must exercise greater control over the specialization s by insisting that it only change the elements of A by a small distance.

We let $\delta = \max\{|x - s(x)| : x \in A\}$. The first thing to note is that choosing δ sufficiently small allows for a canonical choice of $s(u')$ which is itself arbitrarily close to u'. This follows from the fact that a polynomial's roots depend continuously on its coefficients, in the following sense: for a fixed polynomial P and $\varepsilon > 0$, if Q is another polynomial whose coefficients are sufficiently close to the corresponding coefficients of P, then each root of Q is within ε of some root of P. If we choose ε sufficiently small, then each root of Q will be closer to some root of P than any other, and this establishes a bijective correspondence between the roots of P and Q, counting multiplicity [9, pg. 491].

Then let $p(x)$ be the minimal polynomial of u' with coefficients in $\mathbb{Q}[A]$. By applying a specialization with a sufficiently small δ, we can make the coefficients of $p^s(x)$ close to those of $p(x)$, and the above result then allows us to unambiguously choose $s(u')$ to be the root of $p^s(x)$ closest to u', by Proposition 3.7.

This will allow us to resolve our issues. By making δ and $|u' - s(u')|$ sufficiently small, we can ensure that the denominators of the coefficients of all the polynomials in $\mathcal{P}$, which are themselves polynomials in $A \cup \{u'\}$, are moved by a small enough distance that they remain nonzero. Then these coefficients lie in the localization $\mathbb{Q}[A, u']_{\ker(s)}$, and so we can extend s to act on them. This gives us the isomorphism φ^s between f^s and $(f^\sigma)^s$, as required.

The last step is to show that $f^s = f$. The coefficients of f lie in $\mathbb{Q}(a_1, \ldots, a_n, u)$, and s fixes the a_i by design, so we just need $s(u) = u$. We can't choose the value of $s(u)$ at this point, as s was determined by our choice of $s(u')$. However, u is given by a rational function in $A \cup \{u'\}$, and so again by choosing δ and $|u' - s(u')|$ sufficiently small, we can make $s(u)$ close to u. On the other hand, if $q(x)$ is the minimal polynomial of u with coefficients in $\mathbb{Q}[a_1, \ldots, a_n]$, then $s(u)$ must be a root of $q^s(x)$, which is just $q(x)$. By making $s(u)$ closer to u than any other root of q, we can conclude that it *is* u, as required.

In summary, we must choose δ small enough that

- $|u' - s(u')|$ is also small enough to make the next two points work;
- The coefficients of the polynomials in $\mathcal{P}$ lie in $\mathbb{Q}[A, u']_{\ker(s)}$;
- $s(u)$ is closer to u than any other root of $q(x)$.

and these are finitely many constraints, all of which can be accomplished with continuity arguments.

Theorem 3.5 is proved.

3.5 (2) $\Longrightarrow$ (1)

Once we have Theorem 3.5, Belyi's theorem falls out quite nicely from our results on monodromy.

Proposition 3.8. *Let $f : S \to \mathbb{P}^1$ be a meromorphic function ramified over three points. Then the orbit of f under the action of $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$ contains finitely many isomorphism classes.*

Proof. By composing with an appropriate Möbius transformation, we can assume that f ramifies over $\{0, 1, \infty\}$. By Proposition 3.2, f^σ also ramifies over $\{0^\sigma, 1^\sigma, \infty^\sigma\} = \{0, 1, \infty\}$ for all $\sigma \in \mathrm{Gal}(\mathbb{C}/\mathbb{Q})$, and has the same degree.

Now, up to isomorphism, there are only finitely many coverings of a given degree d ramifying over a fixed set of points. This is because, by Theorem 2.13, a covering ramifying over a given set of points is determined up to isomorphism by its monodromy, a homomorphism $\pi(\mathbb{P}^1 \backslash \{0, 1, \infty\}, z_0) \to S_d$. Since this fundamental group is finitely generated [11, pg. 91], there are finitely many of these. So the Galois orbit of f can only contain finitely many coverings. □

With this, we apply Theorem 3.5, and Belyi's theorem is proved. It is informative to note that the proof only really depends on the points of ramification being algebraic, such that there is a finite set of points over which the coverings in the Galois orbit can ramify. The significance of the ramification over 3 points comes from the fact that, by an appropriate Möbius transformation, we can make those points algebraic (for example, $\{0, 1, \infty\}$) whereas there is no guarantee we could do that with 4 or more points.

3.6 Aside: Obvious vs. Non-Obvious

One of the quirks of the discussion surrounding Belyi's theorem is the way its directions are described. That a surface defined over $\overline{\mathbb{Q}}$ admits a covering of the sphere ramified over 3 points is sometime referred to as the "difficult" [8] part of the theorem, even though, as we have seen, the proof is remarkably simple and uses very little machinery beyond the definitions of the concepts involved.

Meanwhile, the reverse direction, that such a covering of the sphere must be definable over $\overline{\mathbb{Q}}$, is sometimes designated as the "obvious" part, which seems at first glance like an even stranger designation. A priori, it is quite surprising that the ramification behavior of meromorphic functions on a surface should tell us about the field of definition. And the proof is certainly more involved than that in the other direction.

However, the historical context backs up these designations. The only part of Belyi's theorem that was actually proven by Belyi was the "difficult" part, while the other direction was known to Grothendieck a few years earlier. Observing Grothendieck's reactions to the two directions of the result, as described in *Esquisse d'un Programme*, sheds some light on why one should be regarded as "obvious" and the other not.

To Grothendieck, the first outstanding surprise of the theory is the fact (described below) that a graph drawn on a topological surface defines a covering of the sphere ramified over 3 points, which in turn gives the surface a canonical structure as a Riemann surface. Next to this, the relevant half of Belyi's theorem merits only a single sentence:

> "Even better, as the complex projective line is defined over the absolute base field $\mathbb{Q}$, as are the admitted points of ramification, the algebraic curves we obtain are defined not only over $\mathbb{C}$, but over the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$ in $\mathbb{C}$." [6, p. 253]

Though it seems technical when proceeding from first principles, this half of the theorem fits very cleanly into the pattern behind much of Grothendieck's work. The machinery of algebraic geometry that he helped develop allows concepts related to coverings and ramification to be defined in purely algebraic ways; then, when their basic parameters lie in $\mathbb{Q}$, it is perfectly natural that the result should be defined over its algebraic closure.

Grothendieck's question is then: now that we have a magical way of pulling Riemann surfaces defined over $\overline{\mathbb{Q}}$ out of embedded graphs, do we get every such surface in this way? This is precisely what the other direction of Belyi's theorem, the one proved by Belyi, answers. Looking at the theorem from this perspective makes Grothendieck's view of its deep significance a bit clearer, as well as the "disconcerting simplicity" of its proof—about which he famously said

> "never, without a doubt, was such a deep and disconcerting result proved in so few lines!"[2]

4 Correspondences

Having proven Belyi's theorem, we put it on the back burner and move on to the other core idea of the theory of dessins d'enfants: some peculiar bijective correspondences between objects that seem unrelated at first glance. In this section, we first describe the three main classes of objects under consideration: dessins d'enfants, constellations, and coverings of the sphere, the latter of which connects us back to Belyi's theorem. Then, we describe the correspondences that allow one to get from one of these objects to another, and roughly explain why these are well-defined up to the different notions of isomorphism we associate to these objects.

The treatment we present here is modeled on chapter 2 of [8], to which the reader is referred for further details.

[2]There seems to be some sort of international law requiring this quote to appear in any survey of dessins d'enfants.

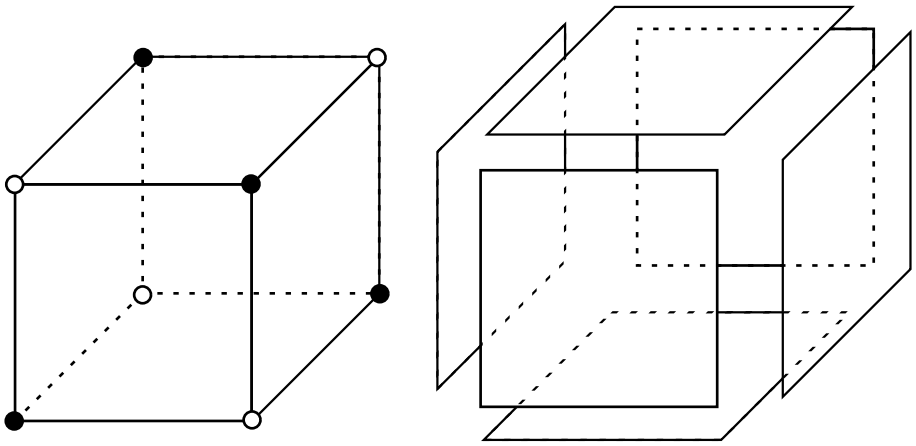


Figure 5: A cube graph embedded on a sphere (that is, a cube), and its faces.

4.1 Object 1: Dessins d'Enfants

The poster child of the rich and varied theory laid out in *Esquisse d'un Programme* is the concept of a **dessin d'enfant**.

Definition 4.1. *A **bicolored graph** is a graph with its vertices colored black and white, such that no two vertices of the same color are adjacent. A **dessin d'enfant** is a bicolored graph embedded on a (topological) orientable surface.*

The term "dessin d'enfant" means "child's drawing", which might say something about Grothendieck's opinion of graph theory.

From a rigorous perspective, this definition requires some careful consideration, because it contains the loaded word "embedded". This refers to a representation of the graph on the surface, where [8, p. 28]:

- the vertices are represented by distinct points of the surface;
- the edges are represented by curves on the surface, whose endpoints are the appropriate vertices, and which do not intersect anywhere else;
- each connected component of the complement of the graph in the surface is homeomorphic to an open disk; these components are the **faces** of the embedded graph.

This last condition can most easily be seen by considering how the graph given by the edges of a regular polyhedron embeds into the polyhedron itself, which is topologically equivalent to a sphere. The resulting faces are what we would normally call the faces of the polyhedron. See Figure 5.

While a great deal of formality goes into the definition, one generally knows an embedded graph when one sees it, although the condition on faces must be handled with care. An embedding of a graph on a surface cannot just be

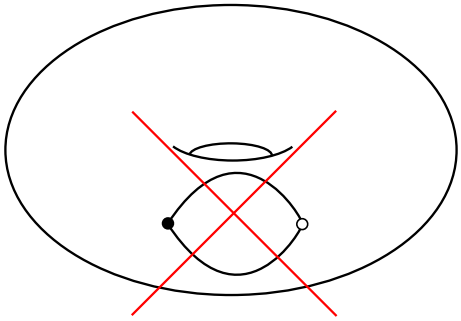


Figure 6: This is not a valid dessin d'enfant. Removing the graph produces two connected components; while one is homeomorphic to a disk, the other has a handle in it.

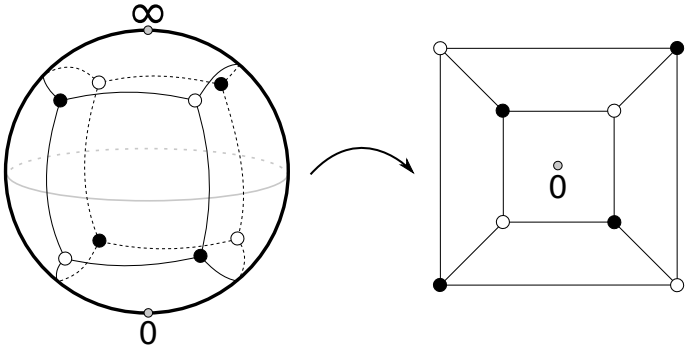


Figure 7: Mapping the cube graph from the sphere to the plane. The top face on the left, which contains ∞, becomes the outside face on the right.

copied over to a surface of higher genus, as this will introduce handles into one or more faces, such that they are no longer homeomorphic to disks.

4.1.1 Depicting Dessins

Since it is difficult to represent compact surfaces on paper, depicting embedded graphs requires some workarounds. We approach the situations of the sphere and surfaces of genus ≥ 1 separately.

Given a graph embedded on the sphere, we can pick any point outside of the graph, call it without loss of generality the point at ∞, and use stereographic projection to map the graph onto the plane. See Figure 7.

Thus we can treat graphs on the sphere in effectively the same way as we treat graphs on the plane. However, we must take care in considering the faces of our graph, as the face containing the point at infinity becomes the "outside" face of the graph in the plane.

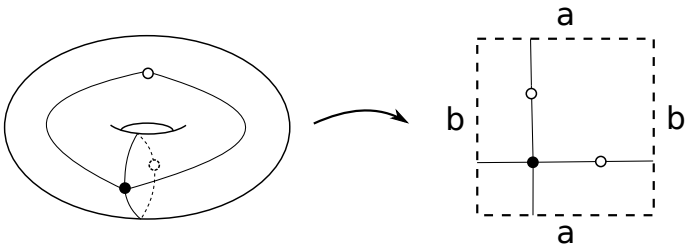


Figure 8: Depicting a dessin embedded on a torus. Note that it has only one face.

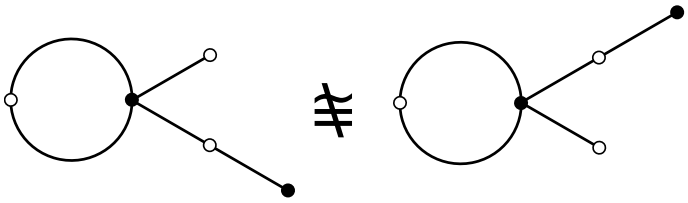


Figure 9: These dessins are not isomorphic.

When working with surfaces of higher genus, we recall that each such surface admits a polygonal presentation [10, Ch. 6]. A topological surface of genus g can be represented by a $2g$-sided polygon with pairs of sides glued together, and we can show the dessin on this polygon instead. An example is shown in Figure 8. This is somewhat harder to draw geometric intuition from, but has the advantage of being flatter than the alternative.

4.1.2 Isomorphism of Dessins

Definition 4.2. *Let D and D' be dessins d'enfants embedded in surfaces S and S', respectively. An **isomorphism** between D and D' is an orientation-preserving homeomorphism $S \to S'$ whose restriction to D is a graph isomorphism which preserves vertex color.*

This definition reflects what is important about dessins: the structures of the abstract graphs as well as their embedding into the ambient space. We note in particular the requirement that an isomorphism be orientation-preserving. We insisted that the surface hosting a dessin be oriented, both because we will be interested in the cyclic ordering of edges emanating from a vertex (which requires orientation to define) and because we will want to give these surfaces Riemann surface structures (which must be oriented).

This means, for instance, that reflecting a dessin around some line on the surface need not produce an isomorphic dessin. See Figure 9.

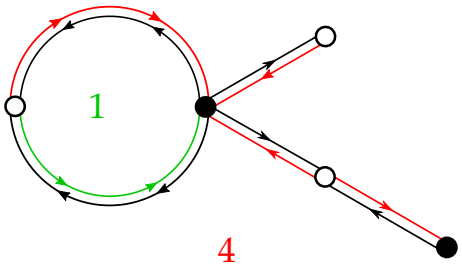


Figure 10: Counting the degrees of faces. Each edge here is doubled to account for the two faces it borders (though they may be the same face) and oriented in the direction associated with a counterclockwise walk around the boundary of that face. We then count only the edges which go from a white to a black vertex. Note that the walk around the outer face appears to be clockwise in this representation, but goes counterclockwise when it is viewed on the sphere.

4.1.3 Degrees of Dessins

In discussing dessins d'enfants, we will often want to talk about faces in the same way we talk about vertices, for reasons that should become clear throughout this section. The most basic concept we want to apply to both vertices and faces is that of degree.

Recall that the **degree** of a vertex is the number of edges incident to it. It would seem natural to define the degree of a face to be the number of edges in its boundary, but the bicoloration complicates this somewhat.

Note that, since every edge is incident to one black vertex and one white vertex, the sum of the degrees of the black vertices, as well as the sum of the degrees of the white vertices, will be the total number of edges. However, edges can border up to two faces, so if we defined this degree as a straightforward count of edges, the sum of the degrees of the faces would generally exceed the total number of edges.

Instead, imagine walking from vertex to vertex counterclockwise around the boundary of a face. This path will hit an equal number of black and white vertices (though potentially landing on some vertices twice) and thus will take an even number of steps. We define the degree of a face to be half of this number. This falls in line with the behavior described above: if we walk such a path for every face, each edge will be used exactly twice, and so the sum of the degrees is the number of edges.

A more concrete way of counting the face degree using this method is to, in going counterclockwise around the face, only count those edges along which we travel from a white vertex to a black vertex. Because the vertices alternate, this will count every other edge and give us the degree. This process is illustrated in Figure 10.

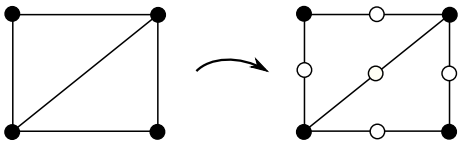


Figure 11: Converting an ordinary graph into a clean dessin.

4.1.4 Maps and "Hypermaps"

Uncharacteristically, Grothendieck did not consider dessins d'enfants in the greatest possible generality. In *Esquisse d'un Programme*, he only makes reference to **maps**, which are embedded graphs without any coloration.

Given a map with all of its vertices colored black, we turn it into a general dessin as described here by placing a white vertex in the middle of each edge, splitting it into two edges. See Figure 11. In the language of maps, then, edges are replaced by **darts** [8] or **flags** [6], half-edges pointing out of a specific endpoint of the larger edge. Described this way, maps are the special case of dessins d'enfants whose white vertices all have degree 2.

[8], starting from the idea of maps, refers to general bicolored dessins as **hypermaps**. As bicolored graphs are the more natural object of study (which the rest of this section will hopefully make clear), and wishing to avoid confusion with the terms "graph" and "hypergraph" (a completely different notion), we will refer to Grothendieck's maps as **clean dessins** (following [3]). In discussing the correspondences below, we will remark on the place clean dessins occupy in other contexts.

4.1.5 Passports

One useful description of the coarse structure of a dessin d'enfant is given by its **passport**. Letting n denote the number of edges of the dessin, this is a triple of partitions of n, given by the unordered collections of the degrees of the black vertices, white vertices, and faces (respectively).

This loses information; one can have nonisomorphic dessins, which may not even be isomorphic as graphs, with the same passport.

4.1.6 Dualing Dessins

A standard notion in the theory of embedded graphs is that of a **dual**, in which one exchanges faces and vertices. With a slight modification to account for the two different types of vertices, we can define a similar notion for dessins.

Given a dessin d'enfant, we construct its dual by placing a black vertex within each face, then connecting that black vertex to each white vertex which appears along the boundary of that face. As with defining the degree of a face, we interpret the boundary of a face as a walk along the incident edges, which

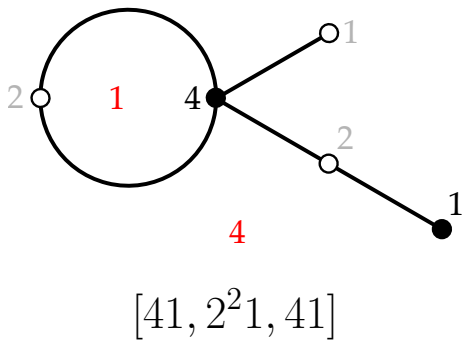


Figure 12: A plane dessin, its vertices and faces (red) labeled with their degrees. From this information, we obtain its passport.

means that white vertices may appear more than once. This requires adding multiple edges.

This is a different notion from the traditional definition of the dual graph [8, pg. 52], which removes all of the original vertices and draws edges between faces instead. However, the two definitions agree within the framework of clean dessins: the (dessin) dual of the clean dessin associated to a map is the same as the clean dessin associated to the (traditional) dual of the map. The reader is invited to verify this for themselves.

4.2 Object 2: Constellations and Permutations

The centerpiece of the theory of dessins d'enfants is a correspondence between dessins and ramified coverings. However, to bridge the two concepts, it is helpful to introduce the intermediate tool of constellations. We will show the equivalence between dessins and coverings by showing that both can be built up from simple pairs of permutations.

Definition 4.3 ([8]). *A **3-constellation**[3] is a triple of permutations $\sigma_0, \sigma_1, \sigma_\infty \in S_n$ such that:*

- *the group generated by these permutations acts transitively on $\{1, \ldots, n\}$, and*
- $\sigma_0 \sigma_1 \sigma_\infty = \mathrm{Id}$.

Of course, this second requirement means that the constellation is determined by the permutations σ_0 and σ_1, and we can recover σ_∞ as $\sigma_1^{-1}\sigma_0^{-1}$. The transitive group generated by the permutations is the **cartographic group** of the constellation.

[3]This definition can be extended to k-constellations consisting of k permutations, but we will not have cause to consider $k \neq 3$ in this paper. Hereafter, all constellations will be 3-constellations.

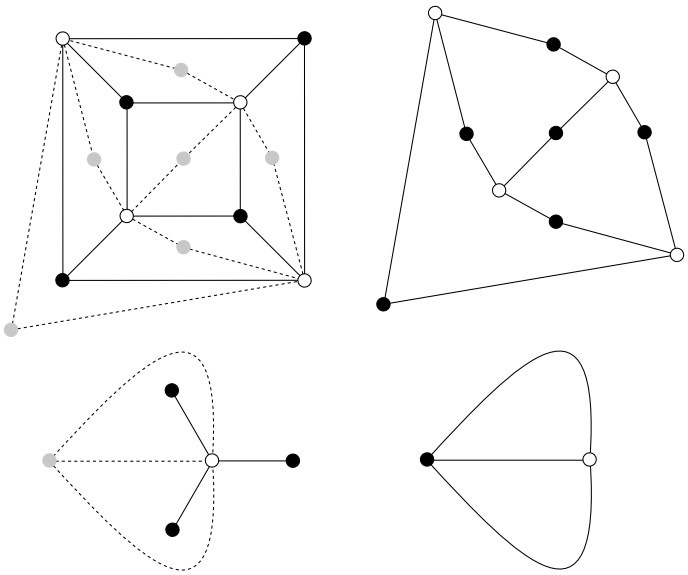


Figure 13: Taking the dual of the cube graph (top) and 3-star (bottom). Note the multiple edges introduced in the latter case.

4.2.1 Isomorphism of Constellations

Intuitively, we should want an isomorphism between two constellations to arise from relabelling the set being permuted. Formally, this is defined as follows:

Definition 4.4. *Let $(\sigma_0, \sigma_1, \sigma_\infty)$ and $(\sigma'_0, \sigma'_1, \sigma'_\infty)$ be constellations permuting sets E and E', respectively. An **isomorphism of constellations** is a bijection $\varphi : E \to E'$ such that $\varphi \circ \sigma_i \circ \varphi^{-1} = \sigma'_i$ for $i = 0, 1, \infty$.*

That is, the permutations which make up the constellation must be simultaneously conjugate.

4.2.2 Passports

One useful description of the coarse structure of a constellation is given by its **passport**. Letting n denote the number of letters permuted by the constellation, this is a triple of partitions of n, given by the cycle types of the permutations.

Example. If $\sigma_0 = (1\ 6\ 8\ 4)(2\ 3\ 7\ 5)$ and $\sigma_1 = (1\ 2\ 5\ 3\ 6\ 4\ 8)(7)$, then $\sigma_1^{-1}\sigma_0^{-1} = (1\ 6\ 8\ 3)(2)(4)(5\ 7)$, and so the passport of this constellation is $[4^2, 71, 421^2]$.

4.3 Object 3: Thrice-Ramified Coverings

We now return to the subject of Belyi's theorem: holomorphic maps $f : S \to \mathbb{P}^1$ which ramify only over $0, 1, \infty$. This section exists mostly to complete the list of the three types of objects to consider, because we have already described all the relevant properties and definitions of these coverings in sections 2 and 3.

We will take this opportunity to say a bit more here about the implications of Belyi's theorem. As stated, the theorem says that these coverings can be defined by polynomials with coefficients in $\overline{\mathbb{Q}}$. However, since there are only finitely many algebraic coefficients defining a covering, they will belong to a more specific finite extension of $\mathbb{Q}$: a **number field**. We say this is a **field of definition** for the covering. For example, the elliptic curve $y^2z = x(x - z)(x - \sqrt{2}z)$ and Belyi map $-4(c_0^2 - 1)/(c_0^2 - 2)^2$ seen in the proof of Belyi's theorem has $\mathbb{Q}(\sqrt{2})$ as a field of definition.

However, as mentioned in the introduction to section 3, this is not well-defined on isomorphism classes: a covering can be defined by many different collections of polynomials, and their coefficients may lie in different fields entirely. To express this number-theoretic information in a way intrinsic to the surface, we introduce the related **field of moduli**.

Recall that the proof of the $(2) \Rightarrow (1)$ direction of Belyi's theorem proceeded by showing that the isomorphism classes of covers ramified over $0, 1, \infty$ fall into finite orbits under the action of the Galois group $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$.

To a covering, we can associate its stabilizer G under this action, and—because the orbit is finite—G has finite index in $\mathrm{Gal}(\mathbb{C}/\mathbb{Q})$. We define the field of moduli of our covering to be the fixed field K of this subgroup: the field of elements unchanged by any element of G. Then the Galois correspondence implies that K is a finite extension of $\mathbb{Q}$.

We can get a stronger result than just finiteness by noting that the field of moduli is contained in every field of definition. If K' is a field of definition for f, then any $\sigma \in \mathrm{Gal}(\mathbb{C}/K')$ fixes f, because it doesn't move any of the coefficients. Thus K must be fixed by $\mathrm{Gal}(\mathbb{C}/K')$, which implies (again by basic properties of the Galois correspondence) that $K \subset K'$. So the field of moduli is, in a sense, the smallest number field associated to an isomorphism class of coverings. A natural question is whether the field of moduli is itself a field of definition, as this would imply there is a realization of the covering over the least complicated extension of $\mathbb{Q}$ possible. While this is true in many cases, it isn't always [8, pg. 122].

For the moment, we simply note that Belyi's theorem shows that our covering comes with some hidden number-theoretic information in terms of the field of moduli, on top of its structure as a Riemann surface or an algebraic curve. When this information is carried through the correspondences we describe here, it produces many surprising patterns and interesting questions.

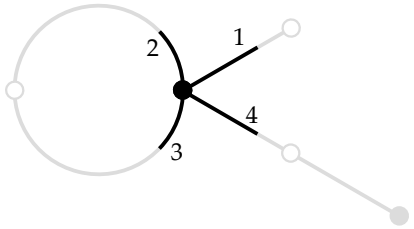


Figure 14: The ordering of edges around a vertex.

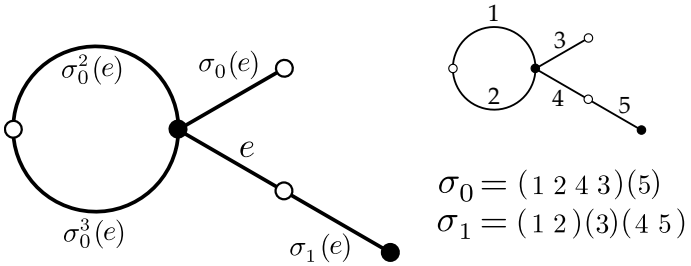


Figure 15: Left: The effect of σ_0 on edges. Right: the permutations σ_0 and σ_1.

4.4 Moving Between Dessins and Permutations

4.4.1 From a Dessin

We insisted above that a dessin d'enfant be embedded in an orientable surface, and now we can use that to succinctly describe it. Looking at a small neighborhood of a vertex, the orientation gives us a counterclockwise ordering of the incident edges (Figure 14).

The bicoloration means that every edge is incident to one black vertex and one white vertex. Then we define a permutation σ_0 on the set of edges, which sends each edge to the next edge in the counterclockwise order around its black vertex. Similarly, we define a permutation σ_1 which sends each edge to the next one in counterclockwise order around its white vertex. (This is unfortunately the reverse of the roles assigned to the black and white vertices in [3]. The same issue will arise in section **??**.) These permutations are depicted in Figure 15.

Now we claim that the group generated by σ_0 and σ_1 is transitive, which means that $(\sigma_0, \sigma_1, \sigma_1^{-1}\sigma_0^{-1})$ is a constellation. This follows from the connectedness of the graph. Between any two edges, we can construct a sequence of adjacent edges, each of which can be obtained from the previous one by a sequence of rotations around the vertices in the path.

Finally, this correspondence is well-defined. Suppose $\varphi : D \to D'$ is an isomorphism of dessins; in an abuse of notation, we use φ to refer both to the underlying homeomorphism and the induced map on edges. Then since φ is an orientation-preserving homeomorphism, it also preserves the cyclic or-

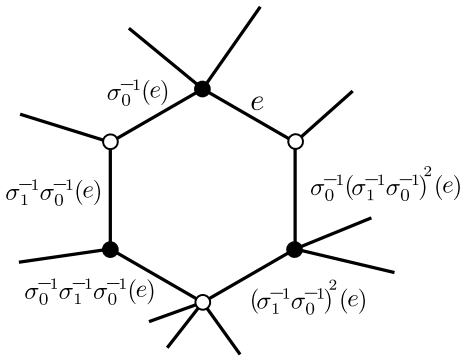


Figure 16: The permutation $\sigma_\infty := \sigma_1^{-1}\sigma_0^{-1}$.

dering of the edges around a given vertex. If $(\sigma_0, \sigma_1, \sigma_\infty)$ and $(\sigma_0', \sigma_1', \sigma_\infty')$ are the constellations associated to D and D', then this is the same as saying that $\varphi \circ \sigma_i \circ \varphi^{-1} = \sigma_i'$ $(i = 0, 1)$, which means that the constellations are isomorphic.

We can now examine how various aspects of a dessin manifest themselves in its constellation. Each black vertex corresponds to a cycle of σ_0, specfically the cycle of edges incident to that vertex. The length of the cycle is the degree of the vertex. Similarly, white vertices correspond to cycles of σ_1.

$\sigma_\infty = \sigma_1^{-1}\sigma_0^{-1}$ has a similar, if slightly less immediate, significance. Just as σ_0 and σ_1 rotate counterclockwise around black and white vertices, σ_∞ rotates counterclockwise around faces, skipping every other edge, as Figure 16 illustrates.

Thus the faces of the dessin correspond to cycles of σ_∞, and the lengths of the cycles are the degrees of the corresponding faces. From this we can see that the passport of a dessin is the same as the passport of its constellation. More importantly, this allows us to read off the genus of the surface our dessin is embedded in.

Proposition 4.5. *Suppose a dessin d'enfant with n edges is embedded on a surface of genus g and has associated constellation $(\sigma_0, \sigma_1, \sigma_\infty)$. Let $c(\sigma_i)$ denote the number of cycles of σ_i. Then*

$$2 - 2g = c(\sigma_0) + c(\sigma_1) + c(\sigma_\infty) - n$$

Proof. This is a direct consequence of Euler's formula [10, p.178–179]: (#vertices)$-$(#edges) $+$ (#faces) $= 2 - 2g$. By the reasoning above, our dessin has $c(\sigma_0)$ black vertices, $c(\sigma_1)$ white vertices, n edges, and $c(\sigma_\infty)$ faces, which gives the result. □

4.4.2 From a Constellation

As shown above, the faces of a dessin are described by the permutation σ_∞. Thus, given a constellation, we can find a dessin associated to it by first using σ_∞ to determine its faces in isolation and then gluing those faces together to

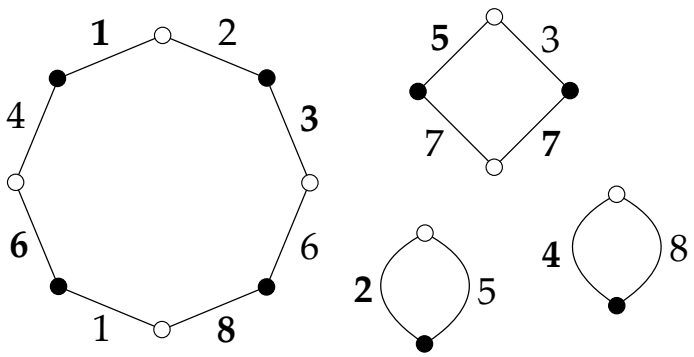
create a compact surface with the graph embedded in it. This is best illustrated by example.

Example. We start with the constellation used in example 3.2.2:

$$\sigma_0 = (1\ 6\ 8\ 4)(2\ 3\ 7\ 5)$$
$$\sigma_1 = (1\ 2\ 5\ 3\ 6\ 4\ 8)(7)$$
$$\sigma_\infty = (1\ 6\ 8\ 3)(2)(4)(5\ 7)$$

First, although it is not necessary to check this ahead of time, we can use Proposition 4.5 to get a preview of the Euler characteristic of the surface in which we will embed our dessin: $2 + 2 + 4 - 8 = 0$, which corresponds to genus 1, so our dessin will be embedded in a torus.

Now we construct the 4 faces of our dessin, recalling that the cycles of σ_∞ only give us every other edge (these edges are bolded in the diagram below). We can obtain the next edge in counterclockwise order by applying σ_0^{-1} to each of the edges in the cycles of σ_∞. For instance, the edge after 1 is $\sigma_0^{-1}(1) = 4$.



In this arrangement, every edge appears twice: once from applying σ_∞ (the bolded edges) and once from applying σ_0^{-1} to some bolded edge. By Proposition 6.4(b) of [10], the quotient space obtained from the faces by identifying these pairs of edges together, orienting them according to the colors of the vertices, is a compact surface. In Figure 17, we show how the polygonal presentation we start with can be maneuvered into a form which is more clearly embedded on the torus. A general description of the type of cutting and pasting used to assemble surfaces can be found in Chapter 6 of [10].

At this point, even before we've done anything exciting using Riemann surfaces, a neat little consequence falls out:

Corollary 4.6. *Suppose $(\sigma_0, \sigma_1, \sigma_1^{-1}\sigma_0^{-1})$ is a constellation. Letting c denote the number of cycles of a permutation, we have*

$$c(\sigma_1^{-1}\sigma_0^{-1}) \leq n + 2 - c(\sigma_0) - c(\sigma_1)$$

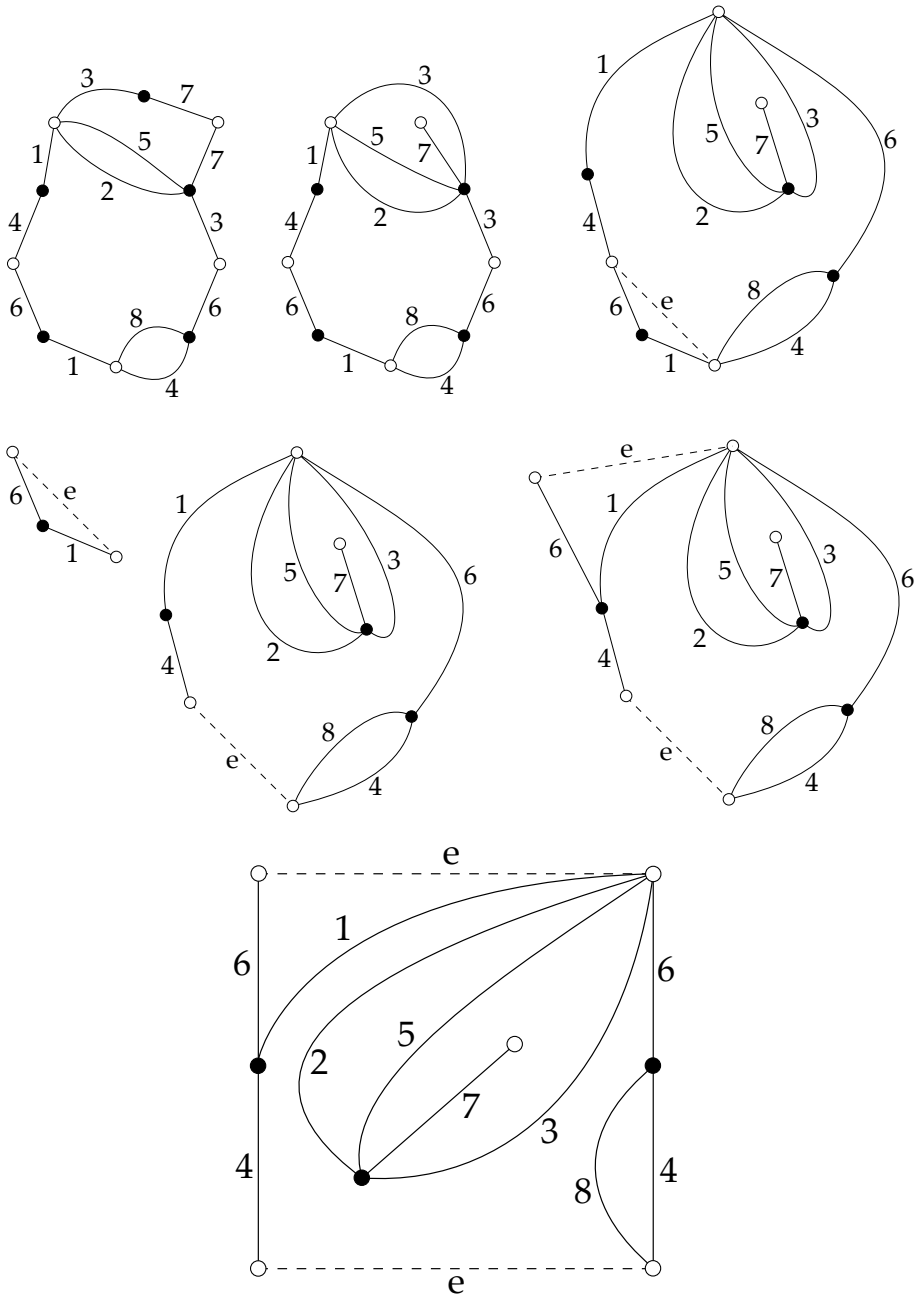


Figure 17: Transforming the above collection of faces into a representation of a dessin on a torus: collapsing the adjacent 7s inward, collapsing the adjacent 3s inward, cutting along the line e, reattaching at edge 1, and finally formatting the whole thing in a more sensible-looking representation.

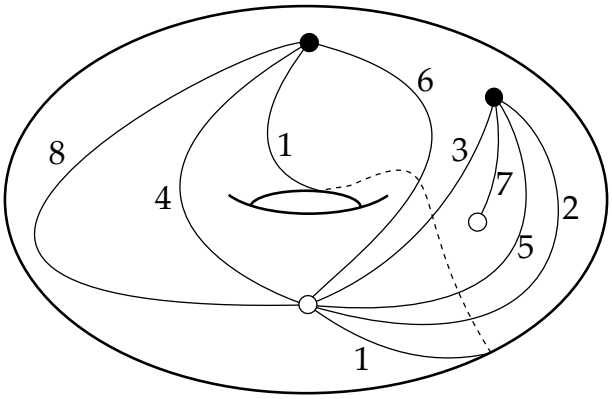


Figure 18: The result of Figure 17 on the more familiar representation of a torus. Observe that the associated permutations are $\sigma_0 = (1\ 6\ 8\ 4)(2\ 3\ 7\ 5)$ and $\sigma_1 = (1\ 2\ 5\ 3\ 6\ 4\ 8)(7)$, the ones we started with.

This follows from Proposition 4.5 and our just-established fact that every constellation corresponds to a dessin. Thus by taking a detour through topology, we have shown a purely combinatorial fact relating the cycle decompositions of two permutations to the cycle decomposition of their product.

4.4.3 Clean Dessins

Since a clean dessin is characterized by all of its white vertices having degree 2, the associated constellations are characterized by all of the cycles of σ_1 having length 2—that is, σ_1 must be an involution without fixed points. If we write out the cycle decomposition of σ_1, each 2-cycle represents a pair of flags which constitute a single edge.

4.5 Moving Between Permutations and Coverings

4.5.1 From a Covering

Suppose $f : S \to \mathbb{P}^1$ is a holomorphic map ramified over $0, 1, \infty$. Then we can obtain a constellation from f using its monodromy.

Identifying $\mathbb{P}^1$ with the sphere and choosing some point $a \neq 0, 1, \infty$, the fundamental group $\pi_1(\mathbb{P}^1 \backslash \{0, 1, \infty\}, a)$ admits a direct description in terms of the three points [11, p. 91]. If we let γ_0, γ_1, and γ_∞ be loops going around their respective points, then it turns out the only relation between their homotopy classes is

$$[\gamma_0][\gamma_1][\gamma_\infty] = 1$$

which can be visualized as follows: a loop representing the left side goes around all 3 points, and can be pulled around the back of the sphere and contracted to a point. See Figure 19.

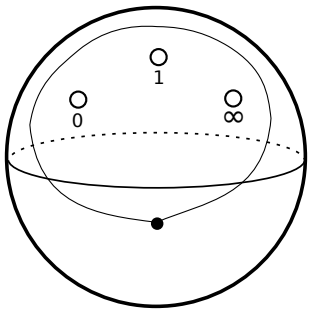


Figure 19: The product $[\gamma_0][\gamma_1][\gamma_\infty]$. Imagine pulling it around the back of the sphere.

Once we know this, we see that the monodromy of this covering is given by three permutations $\sigma_0, \sigma_1, \sigma_\infty$ of $f^{-1}(a)$, corresponding to the generators $[\gamma_0], [\gamma_1], [\gamma_\infty]$, such that $\sigma_0\sigma_1\sigma_\infty = \mathrm{Id}$. Since the monodromy group of a covering must also be transitive, this is exactly a constellation.

Finally, we claim that isomorphic coverings have the same monodromy. Suppose $f : S \to \mathbb{P}^1$ and $f' : S' \to \mathbb{P}^1$ are ramified coverings, $\varphi : S \to S'$ is an isomorphism between them, and γ is a loop in $\mathbb{P}^1$ based at a. Given $x \in f^{-1}(\{a\})$, let $\widetilde{\gamma}$ be a lift of γ starting at x, and let $\widetilde{\gamma}'$ be a lift of γ' starting at $\varphi(x)$. Then

$$f' \circ \widetilde{\gamma}' = \gamma = f \circ \widetilde{\gamma} = f' \circ \varphi \circ \widetilde{\gamma}.$$

Since $\widetilde{\gamma}'$ and $\varphi \circ \widetilde{\gamma}$ are both lifts of γ along f' starting at $\varphi(x)$, they must be identical, and

$$\varphi(x) \cdot [\gamma] = \widetilde{\gamma}'(1) = \varphi(\widetilde{\gamma}(1)) = \varphi(x \cdot [\gamma])$$

Thus φ preserves the monodromy action, and our translation from covers to constellations is well-defined.

4.5.2 From a Permutation

To invert the previous correspondence, given permutations $\sigma_0, \sigma_1 \in S_d$ generating a transitive subgroup, we must construct a covering whose monodromy $\pi_1(\mathbb{P}^1 \backslash \{0, 1, \infty\}, a) \to S_d$ sends $[\gamma_0] \to \sigma_0, [\gamma_1] \to \sigma_1$. Fortunately, we had an entire section about this, back in 2.3.2: Theorem 2.13 shows that there must exist a unique covering $f : X \to \mathbb{P}^1$ with this property. Additionally, since monodromy is only defined up to conjugation, this correspondence is almost trivially defined on isomorphism classes of constellations.

5 Conclusion

We have seen a variety of unexpected connections between different objects, and the natural connection is: where does one go from here?

The most significant goal of this theory is to find ways to describe the group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$—an object fundamental to many questions of number theory—in combinatorial terms. The action of this group on coverings defined over $\overline{\mathbb{Q}}$ gives, by the above correspondences, an action on dessins. On the face of it, there's no reason to assume that this action should manifest itself in any visibly graph-theoretic way, but it does, to a certain extent. One question which is still actively being considered is that of finding Galois invariants: combinatorial properties of dessins which are preserved by the Galois action.

However, as exciting as the big picture of understanding $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is, it's also worthwhile to take stock of the beautiful correspondences shown above. We can simply write down a pair of permutations with transitive action, and out of that pair springs a surface, a graph on that surface, a complex structure on that surface and a holomorphic map to $\mathbb{P}^1$, and a number field. It is an archetypical example of the unity of mathematics in action.

Acknowledgements

My sincerest thanks go to Jim Morrow. Not only was he a helpful mentor throughout the project, offering up fascinating tidbits of background and history and helping me find just the right source on Riemann surfaces, but he also introduced me to the topic of dessins d'enfants, which is a fascinating simulacrum of our academic relationship: I was interested in algebraic number theory, his area of expertise is in complex manifolds, the only work we had done together was on graphs, and somehow he pulled up a subject which unites algebraic number theory with complex manifolds through graphs.

Additional thanks go to friends and family, who put up with me talking incessantly about this thesis.

References

[1] Atiyah, Michael and Ian MacDonald. *Introduction to Commutative Algebra*. Westview Press, 2016.

[2] Bourbaki, Nicolas. *Algèbre, Chapitre 4 à 7*. Springer, 2007. Web.

[3] Girondo, Ernesto and Gabino González-Diez. *Introduction to Compact Riemann Surfaces and Dessins d'Enfants*. Cambridge University Press, 2012.

[4] González-Diez, Gabino. "Variations on Belyi's theorem". *The Quarterly Journal of Mathematics*, vol. 57, no. 3, 2006, pp. 339–354. doi: 10.1093/qmath/hai021. Accessed 17 Apr. 2017.

[5] Griffiths, Phillip and Joseph Harris. *Principles of Algebraic Geometry*. Wiley, 1978.

[6] Grothendieck, Alexandre. "Sketch of a Programme." *Geometric Galois Actions*, edited and translated by Leila Schneps and Pierre Lochak, vol. 1, Cambridge University Press, 1997, 243–283.

[7] Hungerford, Thomas W. *Algebra*. Springer, 1974.

[8] Lando, Sergei K. and Alexander K. Zvonkin. *Graphs on Surfaces and Their Applications*. Springer, 2004.

[9] Lang, Serge. *Algebra*. 3rd ed., Springer, 2002.

[10] Lee, John. *Introduction to Topological Manifolds*. 2nd ed., Springer, 2011.

[11] Miranda, Rick. *Algebraic Curves and Riemann Surfaces*. American Mathematical Society, 1995.

[12] Zieve, Michael (`https://math.stackexchange.com/users/61116/michael-zieve`). "When branch points lie in $\mathbb{P}^1\left(\overline{\mathbb{Q}}\right)$." Math Stack Exchange. URL: `https://math.stackexchange.com/q/746494`.