Geometric and Combinatorial Group Theory

Martin R. Bridson

1 What Are Combinatorial and Geometric Group Theory?

Groups and geometry are ubiquitous in mathematics, groups because the symmetries (or automorphisms (**??**)) of any mathematical object in any context form a group and geometry because it allows one to think intuitively about abstract problems and to organize families of objects into spaces from which one may gain some global insight.

The purpose of this article is to introduce the reader to the study of infinite, discrete groups. I shall discuss both the combinatorial approach to the subject that held sway for much of the twentieth century and the more geometric perspective that has led to an enormous flowering of the subject in the last twenty years. I hope to convince the reader that the study of groups is a concern for all of mathematics rather than something that belongs particularly to the domain of algebra.

The principal focus of *geometric group theory* is the interaction of geometry/topology and group theory, through group actions and through suitable translations of geometric concepts into group theory. One wants to develop and exploit this interaction for the benefit of both geometry/topology and group theory. And, in keeping with our assertion that groups are important throughout mathematics, one hopes to illuminate and solve problems from elsewhere in mathematics by encoding them as problems in group theory.

Geometric group theory acquired a distinct identity in the late 1980s but many of its principal ideas have their roots in the end of the nineteenth century. At that time, low-dimensional topology and *combinatorial group theory* emerged entwined. Roughly speaking, combinatorial group theory is the study of groups defined in terms of *presentations*, that is, by means of generators and relations. In order to follow the rest of this introduction the reader must first understand what these terms mean. Since their definitions would require an unacceptably long break in the flow of our discussion, I will postpone them to the next section, but I strongly advise the reader who is unfamiliar with the meaning of the expression $\Gamma = \langle a_1, \dots, a_n \mid r_1, \dots, r_m \rangle$ to pause and read that section before continuing with this one.

The rough definition of combinatorial group theory just given misses the point that, like many parts of mathematics, it is a subject defined more by its core problems and its origins than by its fundamental definitions. The initial impetus for the subject came from the description of discrete groups of hyperbolic isometries and, most particularly, the discovery of the fundamental group (**??**) of a manifold (**??**) by Poincaré (**??**) in 1895. The group-theoretic issues that emerged were brought into sharp focus by the work of Tietze and Dehn in the first decade of the twentieth century and drove much of combinatorial group theory for the remainder of the century.

Not all of the epoch-defining problems came from topology: other areas of mathematics threw up fundamental questions, typically of forms such as: Does there exist a group of the following type? Which groups have the following property? What are the subgroups of . . . ? Is the following group infinite? When can one determine the structure of a group from its finite quotients? In the sections that follow I shall attempt to illustrate the mathematical culture associated with questions of this kind, but let me immediately mention some easily stated but difficult classical problems. (1) Let G be a group that is finitely generated and suppose that there is some positive integer n such that $x^n = 1$ for every x in G. Must G be finite? (2) Is there a finitely presented group Γ and a surjective homomorphism $\phi : \Gamma \to \Gamma$ such that $\phi(\gamma) = 1$ for some $\gamma \neq 1$? (3) Does there exist a finitely presented, infinite, simple group (**??**)? (4) Is every countable group isomorphic to a subgroup of a finitely generated group, or even a finitely presented group?

The first of these questions was asked by Burnside in 1902 and the second by Hopf in connection with his study of degree-1 maps between manifolds. I shall present the answers to all four questions (in 5) to illustrate an important aspect of both combinatorial and geometric group theory—one develops techniques that allow the construction of *explicit groups* with prescribed properties. Such constructions are of particular interest when they illustrate the diversity of possible phenomena in other branches of mathematics.

Another kind of question that raises basic issues in combinatorial group theory takes the form: Does there exist an algorithm to determine whether or not a group (or given elements of a group) has such-and-such a property? For example, does there exist an algorithm that can take any finite presentation and decide in a finite number of steps whether or not the group presented is trivial? Questions of this type led to a profound and mutually beneficial interaction between group theory and logic, given full voice by the Higman embedding theorem, which we shall discuss in 6. Moreover, via the conduit of combinatorial group theory, logic has influenced topology as well: one uses group-theoretic constructions to show, for example, that there is no algorithm to determine which pairs of compact triangulated manifolds are homeomorphic in dimensions 4 and above. This shows that certain kinds of classification results that have been obtained in two and three dimensions do not have higher-dimensional analogues.

One might reasonably regard combinatorial group theory as the attempt to develop algebraic techniques to solve the types of questions described above, and in the course of doing so to identify classes of groups that are worthy of particular study. This last point, the question of which groups deserve our attention, is tackled head-on in the final section of this article.

Some of the triumphs of combinatorial group theory are intrinsically combinatorial in nature, but many more have had their true nature revealed by the introduction of geometric techniques in the past twenty years. A fine example of this is the way in which Gromov's insights have connected algorithmic problems in group theory to so-called filling problems in Riemannian geometry. Moreover, the power of geometric group theory is by no means confined to improving the techniques of combinatorial group theory: it naturally leads one to think about many other issues of fundamental importance. For example, it provides a context in which one can illuminate and vastly extend classical RIGIDITY THEOREMS (**??**), such as that of Mostow. The key to applications such as this is the idea that finitely generated groups can usefully be regarded as geometric objects in their own right. This idea has its origins in the work of CAYLEY (**??**) (1878) and Dehn (1905) but its full force was recognized and promoted by Gromov, starting in the 1980s. It is the key idea that underpins the later parts of this article.

2 Presenting Groups

How should one describe a group? An example will illustrate the standard way of doing so and give some idea of why it is often appropriate.

Consider the familiar tiling of the Euclidean plane by equilateral triangles. How might you describe the full group Γ_Δ of symmetries of this tiling, i.e., the rigid motions of the plane that send tiles to tiles? Let us focus on a single tile T and a particular edge e of T, and use this to pick out three symmetries. The first, which we shall call α, is the reflection of the plane in the line that contains e and the other two, β and γ, are the reflections in the lines that join the endpoints of e to the midpoints of the opposite edges in T. With some effort one can convince oneself that every symmetry of the tiling can be obtained by performing these three operations repeatedly in a suitable order. One expresses this by saying that the set $\{\alpha, \beta, \gamma\}$ *generates* the group Γ_Δ.

A further useful observation is that if one performs the operation α twice, the tiling is returned to its original position: that is, $\alpha^2 = 1$. Likewise, $\beta^2 = \gamma^2 = 1$. One can also verify that $(\alpha\beta)^6 = (\alpha\gamma)^6 = (\beta\gamma)^3 = 1$.

It turns out that the group Γ_Δ is completely determined by these facts alone—a statement that we summarize by the notation

$$\Gamma_\Delta = \langle \alpha, \beta, \gamma \mid \alpha^2, \beta^2, \gamma^2, (\alpha\beta)^6, (\alpha\gamma)^6, (\beta\gamma)^3 \rangle.$$

The aim of the rest of this section is to say in more detail what this means.

To begin with, notice that from the facts we are given we can deduce others: for example, bearing in mind that $\beta^2 = \gamma^2 = (\beta\gamma)^3 = 1$, we can show that

$$(\gamma\beta)^3 = (\gamma\beta)^3(\beta\gamma)^3 = 1$$

as well (where the last equality follows after repeatedly canceling pairs of the form $\beta\beta$ or $\gamma\gamma$). We wish to convey the idea that in Γ_Δ there are no relationships between the generators except those that follow from the facts above by this kind of argument.

Now let us try to say this more formally. We define a *set of generators* for a group Γ to be a subset $S \subset \Gamma$ such that every element of Γ is equal to some product of elements of S and their inverses. That is, every element can be written in the form $s_1^{\varepsilon_1} s_2^{\varepsilon_2} \cdots s_n^{\varepsilon_n}$, where each s_i is an element of S and each ε_i is 1 or -1. We then call a product of this kind a *relation* if it is equal to the identity in Γ.

There is an awkward ambiguity here. When we talk about "the product" of some elements of Γ, it sounds as though we are referring to another element of Γ, but we certainly did not mean this at the end of the last paragraph: a relation is not the identity element of Γ but rather a *string of symbols* such as $ab^{-1}a^{-1}bc$ that yields the identity in Γ when you interpret a, b, and c as generators in the set S. In order to be clear about this, it is useful to define another group, known as the *free group* $F(S)$.

For concreteness we shall describe the free group with three generators, taking our set S to be $\{a, b, c\}$. A typical element is a "word" in the elements of S and their inverses, such as the expression $ab^{-1}a^{-1}bc$ considered in the previous paragraph. However, we sometimes regard two words as the same: for instance, $abcc^{-1}ac$ and $abab^{-1}bc$ are the same because they become identical when we cancel out the inverse pairs cc^{-1} and $b^{-1}b$. More formally, we define two such words to be *equivalent* and say that the elements of the free group are the EQUIVALENCE CLASSES (**??**). To multiply words together, we just concatenate them: for instance, the product of ab^{-1} and $bcca$ is $ab^{-1}bcca$, which we can shorten to $acca$. The identity is the "empty word." This is the free group on three generators a, b, c. It should be clear how to generalize it to an arbitrary set S, though we shall continue to discuss the set $S = \{a, b, c\}$.

A more abstract way of characterizing the free group on a, b, and c is to say that it has the following *universal property*: if G is any group and ϕ is any function from $S = \{a, b, c\}$ to G, then there is a unique homomorphism Φ from $F(S)$ to G that takes a to $\phi(a)$, b to $\phi(b)$, and c to $\phi(c)$. Indeed, if we want Φ to have these properties, then our definition is forced upon us: for example, $\Phi(ab^{-1}ca)$ will have to be $\phi(a)\phi(b)^{-1}\phi(c)\phi(a)$, by the definition of a homomorphism. So the uniqueness is obvious. The rough reason that this definition really does give rise to a well-defined homomorphism is that the only equations that are true in $F(S)$ are ones that are true in all groups: in order for Φ not to be a homomorphism, one would need a relation to hold in $F(S)$ that did not hold in G, but this is impossible.

Now let us return to our example Γ_Δ. We would like to prove that it is (isomorphic to) the "freest" group with generators α, β, and γ that satisfies the relations $\alpha^2 = \beta^2 = \gamma^2 = (\alpha\beta)^6 = (\alpha\gamma)^6 = (\beta\gamma)^3 = 1$. But what exactly is this "freest" group that we are claiming is isomorphic to Γ_Δ?

To avoid confusion about the meaning of α, β, and γ (are they elements of Γ_Δ or of the group that we are trying to construct that will turn out to be isomorphic to Γ_Δ?) we shall use the letters a, b, and c when we answer this question. Thus, we are trying to build the "freest" group with generators a, b, and c that satisfies the relations $a^2 = b^2 = c^2 = (ab)^6 = (ac)^6 = (bc)^3 = 1$, which we denote by $G = \langle a, b, c \mid a^2, b^2, c^2, (ab)^6, (ac)^6, (bc)^3 \rangle$.

There are two ways of going about this task. One is to imitate the above discussion of the free group itself, except that now we say that two words are equivalent if you can get from one to the other by inserting or deleting not just inverse pairs but also one of the words a^2, b^2, c^2, $(ab)^6$, $(ac)^6$, or $(bc)^3$. For example, ab^2c is equivalent to ac in this group. G is then defined to be the set of equivalence classes of words with the product coming from concatenation.

A neater way to obtain G is more conceptual and exploits the universal property of the free group. As G is to be generated by a, b, and c, the universal property of the free group $F(S)$ tells us that there will have to be a unique homomorphism Φ from $F(S)$ to G such that $\Phi(a) = a$, $\Phi(b) = b$, and $\Phi(c) = c$. Moreover, we require that all of a^2, b^2, c^2, $(ab)^6$, $(ac)^6$, and $(bc)^3$ must map to the identity element in G. It follows that the kernel of Φ is a normal subgroup of $F(S)$ that contains the set $R = \{a^2, b^2, c^2, (ab)^6, (ac)^6, (bc)^3\}$. Let us write $\langle\langle R \rangle\rangle$ for the smallest normal subgroup of $F(S)$ that contains R (or equivalently the intersection of all normal subgroups of $F(S)$ that contain R). Then there is a surjective homomorphism from the QUOTIENT (**??**) $F(S)/\langle\langle R \rangle\rangle$ to *any* group that is generated by a, b, and c and satisfies the relations $a^2 = b^2 = c^2 = (ab)^6 = (ac)^6 = (bc)^3 = 1$. This quotient itself is the group we are looking for: it is the largest group generated by a, b, and c that satisfies the relations in R.

Our assertion about Γ_Δ is that it is isomorphic to the group $G = \langle a, b, c \mid a^2, b^2, c^2, (ab)^6, (ac)^6, (bc)^3 \rangle$ that we have just described (in two ways). More precisely, the map from $F(S)/\langle\langle R \rangle\rangle$ to Γ_Δ that takes a to α, b to β, and c to γ is an isomorphism.

The above construction is very general. If we are given a group Γ, then a *presentation* of Γ is a set S that generates Γ, together with a set $R \subset F(S)$ of relations, such that Γ is isomorphic to the quotient $F(S)/\langle\langle R \rangle\rangle$. If both S and R are finite sets, one says that the presentation is finite. A group is *finitely presented* if it has a finite presentation.

We can also define presentations in the abstract, without mentioning a group Γ in advance: given any set S and any subset $R \subset F(S)$, we just define $\langle S \mid R \rangle$ to be the group $F(S)/\langle\langle R \rangle\rangle$. This is the "freest" group generated by S that satisfies the relations in R: the only relations that hold in $\langle S \mid R \rangle$ are the ones that can be deduced from the relations R.

A psychological advantage of switching to this more abstract setting is that, whereas previously we began with a group Γ and asked how we might present it, we can now write down group presentations at will, starting with any set S and prescribing a set of words R in the symbols $S^{\pm 1}$. This gives us a very flexible way of constructing a wide variety of groups. We might, for example, use a group presentation to encode a question from elsewhere in mathematics. We could then ask about the properties of the group thus defined, and see what they had to tell us about our original problem.

3 Why Study Finitely Presented Groups?

Groups arise across the whole of mathematics as *groups of automorphisms*. These are maps from an object to itself that preserve all of the defining structure: two examples are the invertible LINEAR MAPS (**??**) from a VECTOR SPACE (**??**) to itself, and the homeomorphisms from a TOPOLOGICAL SPACE (**??**) to itself. Groups encapsulate the essence of symmetry and for this reason demand our attention. We are driven to understand their general nature, identify groups that deserve particular attention, and develop techniques for constructing new groups (from old ones, or from new ideas). And, reversing the process of abstraction, when *given* a group, we want to find concrete instances of it. For example, we might like to realize it as the group of automorphisms of some interesting object, with the aim of illuminating the nature of both the object and the group. (See, for example, the article on representation theory (**??**).)

3.1 Why Present Groups in Terms of Generators and Relations?

The short answer is that this is the form in which groups often "appear in nature." This is particularly true in topology. Before looking at a general result that illustrates this point, let us examine a simple example. Consider the group D of all isometries of $\mathbb{R}$ that are generated by the reflections at the points 0, 1, and 2: that is, the group generated by the three functions α_0, α_1, and α_2, which take x to $-x$, $2-x$, and $4-x$, respectively. You may recognize this group to be the infinite dihedral group, and you may notice that the generator α_2 is superfluous, since it can be generated from α_0 and α_1. But let us close our eyes to these observations as we let a presentation emerge from the action.

To this end, we choose an open interval U with the property that the images of U under the maps in D cover the whole of the real line, say $U = (-\frac{1}{2}, \frac{3}{2})$. Now let us record two pieces of data: the only elements of D (apart from the identity) that fail to move U completely off itself are α_0 and α_1, and, among all products of length at most 3 in those two letters, the only nontrivial ones that act as the identity on $\mathbb{R}$ are α_0^2 and α_1^2. You may like to prove that $\langle \alpha_0, \alpha_1 \mid \alpha_0^2, \alpha_1^2 \rangle$ is a presentation of D.

This is in fact a special case of a general result, which we now state. (The proof of it is somewhat involved.) Let X be a topological space that is both path connected (**??**) and simply connected (**??**), and let Γ be a group of homeomorphisms from X to itself. Then any choice of path-connected open subset $U \subset X$ such that the images of U cover all of X gives rise to a presentation $\Gamma = \langle S \mid R \rangle$, where $S = \{\gamma \in \Gamma \mid \gamma(U) \cap U \neq \emptyset\}$ and R consists of all words $w \in F(S)$ of length at most 3 such that $w = 1$ in Γ. Thus, the identification of a suitable subset U provides one with a presentation of Γ, and the task of a group theorist is to determine the nature of the group from this information.

To see how difficult this task is, you might like to consider the groups

$$G_n = \langle a_1, \dots, a_n \mid a_i^{-1} a_{i+1} a_i a_{i+1}^{-2},\ i = 1, \dots, n \rangle,$$

where we interpret $i+1$ as 1 when $i = n$. One of G_3 and G_4 is trivial and the other is infinite. Can you decide which is which?

To illustrate a more subtle point, let us consider a finitely presented group that we perhaps feel we understand: the group Γ_Δ that we were discussing earlier. If we want to describe this group to a blind friend unfamiliar with the triangular tiling of the plane, what can we say to make her understand the group, or at least convince her that we understand the group?

Our friend might reasonably ask us to list the elements of our group, so we begin to describe them as products (words) in the given generators. But as we begin to do so we hit a problem: we do not want to list any element more than once and in order to avoid redundancy we have to know which pairs of words w_1, w_2 represent the same element of Γ_Δ; equivalently, we must be able to recognize which words $w_1^{-1} w_2$ are relations in the group. Determining which words are relations is called the *word problem* for the group. Even in Γ_Δ this takes some work, and in the groups G_n we quickly find ourselves at a loss.

Note that as well as allowing one to list the elements of the group effectively, a solution to the word problem also allows one to determine the multiplication table, since deciding whether $w_1 w_2 = w_3$ is the same as deciding whether $w_1 w_2 w_3^{-1} = 1$.

3.2 Why *Finitely* Presented Groups?

The packaging of infinite objects into finite amounts of data arises throughout mathematics in the various guises of compactness (**??**). Finite presentation is basically a compactness condition: a group can be finitely presented if and only if it is the fundamental group of a reasonable compact space, as we shall see later.

Another reason for studying finitely presented groups is that the Higman embedding theorem (to be discussed later) allows us to encode questions about arbitrary Turing machines as questions about such groups and their subgroups.

4 The Fundamental Decision Problems

In exploring the geometry and topology of low-dimensional manifolds at the beginning of the twentieth century, Max Dehn saw that many of the problems that he was wrestling with could be "reduced" to questions about finitely presented groups. For example, he gave a simple formula for associating with a knot diagram (**??**) a finite presentation of a group. There was one relation for each crossing in the diagram and he argued that the resulting group would be isomorphic to $\mathbb{Z}$ if and only if the knot was the unknot: that is, if and only if it could be continuously deformed into a circle. It is extremely hard to tell by staring at a knot diagram whether it is actually the unknot, so this seems like a useful reduction until one realizes that it can be just as hard to tell whether a finitely presented group is isomorphic to $\mathbb{Z}$. For example, here is the presentation of $\mathbb{Z}$ that Dehn's recipe associates with one of smallest possible pictures of the unknot, namely a diagram with just four crossings:

$$\langle a_1, a_2, a_3, a_4, a_5 \mid a_1^{-1}a_3a_4^{-1}, a_2a_3^{-1}a_1, a_3a_4^{-1}a_2^{-1}, a_4a_5^{-1}a_4a_3^{-1}\rangle.$$

Thus Dehn's investigations led him to understand how difficult it is to extract information from a group presentation. In particular, he was the first to identify the fundamental role of the word problem, which we alluded to earlier, and he was one of the first to begin to understand that there are fundamental problems associated with the challenge of developing *algorithms* that extract knowledge from well-defined objects such as group presentations. In his famous article of 1912 Dehn writes:

> The general discontinuous group is given by n generators and m relations between them.... Here *there are above all three fundamental problems* whose solution is very difficult and which will not be possible without a penetrating study of the subject.
>
> **1. The identity [word] problem:** An element of the group is given as a product of generators. One is required to give a method whereby it may be decided in a finite number of steps whether this element is the identity or not.
>
> **2. The transformation [conjugacy] problem:** Any two elements S and T of the group are given. A method is sought for deciding the question whether S and T can be transformed into each other, i.e. whether there is an element U of the group satisfying the relation
> $$S = UTU^{-1}.$$
> **3. The isomorphism problem:** Given two groups, one is to decide whether they are isomorphic or not (and further, whether a given correspondence between the generators of one group and elements of the other is an isomorphism or not).

We shall take these problems as the starting point for three lines of enquiry. First, we shall work toward an outline of the proof that all of these problems are, in a strict sense, unsolvable for general finitely presented groups.

The second use that we shall make of Dehn's problems is to hold them up as fundamental measures of complexity for each of the classes of groups that we subsequently encounter. If we can prove, for example, that the isomorphism problem is solvable in one class of groups but not in another, then we will have given genuine substance to previously vague assertions to the effect that the second class is "harder."

Finally, I want to make the point that geometry lies at the heart of the fundamental issues in combinatorial group theory: it may not be immediately obvious, but its implicit presence is nonetheless a fundamental trait of group theory and not something imposed for reasons of taste. To illustrate this point I shall explain how the study of the large-scale geometry of least-area disks in Riemannian manifolds is intimately connected with the study of the complexity of word problems in arbitrary finitely presented groups.

5 New Groups from Old

Suppose that you have two groups, G_1 and G_2, and want to combine them to form a new group. The first method that is taught in a typical course on group theory is to take the Cartesian product $G_1 \times G_2$: a typical element has the form (g, h) with $g \in G_1$ and $h \in G_2$, and the product of (g, h) with (g', h') is defined to be (gg', hh'). The set of elements of the form (g, e) (where e is the identity of G_2) is a copy of G_1 inside $G_1 \times G_2$, and similarly the set of elements of the form (e, h) is a copy of G_2.

These copies have nontrivial relations between their elements: for example, $(e, h)(g, e) = (g, e)(e, h)$. We would now like to take two groups Γ_1 and Γ_2 and combine them in a different way to form a group called the *free product* $\Gamma_1 * \Gamma_2$, which contains copies of Γ_1 and Γ_2 and as few additional relations as possible. That is, we would like there to be embeddings $i_j : \Gamma_j \hookrightarrow \Gamma_1 * \Gamma_2$ so that $i_1(\Gamma_1)$ and $i_2(\Gamma_2)$ generate $\Gamma_1 * \Gamma_2$ but they are not intertwined in any way. This requirement is neatly encapsulated by the following universal property: given any group G and any two homomorphisms $\phi_1 : \Gamma_1 \to G$ and $\phi_2 : \Gamma_2 \to G$, there should be a unique homomorphism $\Phi : \Gamma_1 * \Gamma_2 \to G$ such that $\Phi \circ i_j = \phi_j$ for $j = 1, 2$. (Less formally, Φ behaves like ϕ_1 on the copy of Γ_1 and behaves like ϕ_2 on the copy of Γ_2.)

It is easy to check that this property characterizes $\Gamma_1 * \Gamma_2$ up to isomorphism, but it leaves open the question of whether $\Gamma_1 * \Gamma_2$ actually exists. (These are the standard pros and cons of defining an object by means of a universal property.) In the present setting, existence is easily established using presentations: let $\langle A_1 \mid R_1\rangle$ be a presentation of Γ_1 and let $\langle A_2 \mid R_2\rangle$ be a presentation of Γ_2, with A_1 and A_2 disjoint, and then define $\Gamma_1 * \Gamma_2$ to be $\langle A_1 \sqcup A_2 \mid R_1 \sqcup R_2\rangle$ (where $\sqcup$ denotes a union of disjoint sets).

More intuitively, one can define $\Gamma_1 * \Gamma_2$ to be the set of alternating sequences $a_1 b_1 \cdots a_n b_n$ with each a_i belonging to Γ_1 and each b_j belonging to Γ_2, with the extra condition that none of the a_i and b_j equals the identity, except possibly a_1 or b_n. The group operations in Γ_1 and Γ_2 extend to this set in an obvious way: for example, $(a_1 b_1 a_2)(a_1' b_1') = a_1 b_1 a_2' b_1'$, where $a_2' = a_2 a_1'$, except that if $a_2 a_1' = 1$ then the product cancels down to $a_1 b_2'$, where $b_2' = b_1 b_1'$.

Free products occur naturally in topology: if one has topological spaces X_1, X_2 with marked points $p_1 \in X_1$, $p_2 \in X_2$, then the fundamental group of the space $X_1 \vee X_2$ obtained from $X_1 \sqcup X_2$ by making the identification $p_1 = p_2$ is the free product of $\pi_1(X_1, p_1)$ and $\pi_1(X_2, p_2)$. The Seifert–van Kampen theorem tells one how to present the fundamental group of a space obtained by gluing X_1 and X_2 along larger subspaces. If the inclusion of the subspaces gives rise to an injection of fundamental groups, then one can express the fundamental group of the resulting space as an *amalgamated free product*, which we now define.

Let Γ_1 and Γ_2 be two groups. If some other group contains copies of Γ_1 and Γ_2, then the intersection of those copies must contain the identity element. The free product $\Gamma_1 * \Gamma_2$ was the freest group we could build that was subject to this minimal constraint. Now we shall insist that the copies of Γ_1 and Γ_2 intersect nontrivially, specify which of their subgroups must lie in the intersection, and build the freest group that satisfies this constraint.

Suppose, then, that A_1 is a subgroup of Γ_1 and that ϕ is an isomorphism from A_1 to a subgroup A_2 of Γ_2. As in the example of the free product, one can define the "freest product that identifies A_1 and A_2" by means of a universal property. Again, one can establish the existence of such a group using presentations: if $\Gamma_1 = \langle S_1 \mid R_1 \rangle$ and $\Gamma_2 = \langle S_2 \mid R_2 \rangle$, the group we seek takes the form

$$\langle S_1 \sqcup S_2 \mid R_1 \sqcup R_1 \sqcup T \rangle.$$

Here, $T = \{u_a v_a^{-1} \mid a \in A_1\}$, where u_a is some word that represents a in (the presentation of) Γ_1 and v_a is a word that represents $\phi(a)$ in Γ_2.

This group is called the *amalgamated free product of* Γ_1 *and* Γ_2 *along* A_1 *and* A_2. It is often described by the casual and ambiguous notation $\Gamma_1 *_{A_1 = A_2} \Gamma_2$, or even $\Gamma_1 *_A \Gamma_2$, where $A \cong A_j$ is an abstract group.

Unlike with free products, it is no longer obvious that the maps $\Gamma_i \to \Gamma_1 *_A \Gamma_2$ implicit in this construction are injective, but they do turn out to be, as was shown by Schreier in 1927.

A related construction of Higman, Neumann, and Neumann in 1949 answers the following question: given a group Γ and an isomorphism $\psi : B_1 \to B_2$ between subgroups of Γ, can one always embed Γ in a bigger group so that ψ becomes the restriction to B_1 of a conjugation?

By now, having seen the idea in the context of both free products and amalgamated free products, the reader may guess how one goes about answering this question: one writes down the presentation of a universal candidate for the desired enveloping group, denoted $\Gamma *_\psi$, and then one sets about proving that the natural map from Γ to $\Gamma *_\psi$ (which takes each word to itself) is injective. Thus, given $\Gamma = \langle A \mid R \rangle$, we introduce a symbol $t \notin A$ (usually called *the stable letter*), we choose for each $b \in B_1$ words $\hat{b}, \tilde{b} \in F(A)$ with $\hat{b} = b$ and $\tilde{b} = \psi(b)$ in Γ, and we define

$$\Gamma *_\psi := \langle A, t \mid R, t\hat{b}t^{-1}\tilde{b}^{-1} \ (b \in B_1) \rangle.$$

This is the freest group we can build from Γ by adjoining a new element t and requiring it to satisfy all the equations we want it to, namely $t\hat{b}t^{-1} = \tilde{b}$ for every $b \in B_1$ (which we can think of as saying that $tbt^{-1} = \psi(b)$). This group is called an *HNN extension* of Γ (after Higman, Neumann, and Neumann).

Now we must show that the natural map from Γ to $\Gamma *_\phi$ is injective. That is, if you take an element γ of Γ and regard it as an element of $\Gamma *_\psi$, you should not be able to use t and the relations in $\Gamma *_\psi$ to cancel γ down to the identity. This is proved with the help of the following more general result known as *Britton's lemma*. Suppose that w is a word in the free group $F(A, t)$. Then the only circumstances under which it can give rise to the identity in the group $\Gamma *_\psi$ are if either it does not involve t and represents the identity in Γ or it involves t but can be simplified in an obvious way by containing a "pinch." A pinch is a subword of the form tbt^{-1}, where b is a word in $F(A)$ that represents an element of B_1 (in which case we can replace it by $\psi(b)$), or one of the form $t^{-1}b't$, where b' represents an element of B_2 (in which case we can replace it by $\psi^{-1}(b')$). Thus, if you are given a word that involves t and contains no pinches, then you know that it cannot be canceled down to the identity.

A similar noncancellation result holds for the amalgamated free product $\Gamma_1 *_{A_1 = A_2} \Gamma_2$. If $g_1, \dots, g_n$ belong to Γ_1 but not to A_1 and $h_1, \dots, h_n$ belong to Γ_2 but not to A_2, then the word $g_1 h_1 g_2 h_2 \cdots g_n h_n$ cannot equal the identity in $\Gamma_1 *_{A_1 = A_2} \Gamma_2$.

These noncancellation results do far more than show that the natural homomorphisms we have been considering are injective: they also demonstrate further aspects of freeness in amalgamated free products and HNN extensions. For example, suppose that in the amalgamated free product $\Gamma_1 *_{A_1 = A_2} \Gamma_2$ we can find an element g of Γ_1 that generates an infinite group which intersects A_1 in the identity and an element h of Γ_2 that does the same for A_2. Then the subgroup of $\Gamma_1 *_{A_1=A_2} \Gamma_2$ generated by g and h is the free group on those two generators. With a little more effort, one can deduce that any finite subgroup of $\Gamma_1 *_{A_1=A_2} \Gamma_2$ has to be conjugate to a subgroup of the obvious copy of either Γ_1 or Γ_2. Similarly, the finite subgroups of $\Gamma *_\psi$ are conjugates of subgroups of Γ. We shall exploit these facts in the constructions that follow.

There are many ways of combining groups that I have not mentioned here. I have chosen to focus on amalgamated free products and HNN extensions partly because they lead to transparent solutions of the basic problems discussed below but more because of their primitive appeal and the way in which they arise naturally in the calculation of fundamental groups. They also mark the beginning of *arboreal group theory*, which we will discuss later. If space allowed, I would go on to describe semidirect and wreath products, which are also indispensable tools of the group theorist.

Before turning to some applications of HNN extensions and amalgamated free products, I want to return to the Burnside problem, which asks if there exist finitely generated infinite groups all of whose elements have a given finite order. This question generated important developments throughout the twentieth century, particularly in Russia. It is appropriate to mention it here because it provides another illustration of the fact that it can be useful to study a universal object in order to solve a general question.

5.1 The Burnside Problem

Given an exponent m, one clarifies the problem at hand by considering the *free Burnside group* $B_{n,m}$ given by the presentation $\langle a_1, \dots, a_n \mid R_m \rangle$, where R_m consists of all mth powers in the free group $F(a_1, \dots, a_n)$. It is clear that $B_{n,m}$ maps onto any group with at most n generators in which every element has order dividing m. Therefore, there exists a finitely generated infinite group with all elements of the same finite order if and only if, for suitable values of n and m, the group $B_{n,m}$ is infinite. Thus, a question that takes the form, Does there exist a group such that ...?, becomes a question about just one group.

Novikov and Adian showed in 1968 that $B_{n,m}$ is infinite when $n \geqslant 2$ and $m \geqslant 667$ is odd. Determining the exact range of values for which $B_{n,m}$ is infinite is an active area of research. Of far greater interest is the open question of whether there exist finitely presented infinite groups that are quotients of $B_{n,m}$. Zelmanov was awarded the Fields Medal for proving that each $B_{n,m}$ has only finitely many finite quotients.

5.2 Every Countable Group Can Be Embedded in a Finitely Generated Group

Given a countable group G we list its elements, $g_0, g_1, g_2, \dots$, taking g_0 to be the identity. We then take a free product of G with an infinite cyclic group $\langle s \rangle \cong \mathbb{Z}$. Let Σ_1 be the set of all elements of $G * \mathbb{Z}$ of the form $s_n = g_n s^n$ with $n \geqslant 1$. Then the subgroup $\langle \Sigma_1 \rangle$ generated by Σ_1 is isomorphic to the free group $F(\Sigma_1)$. Similarly, if we let $\Sigma_2 = \{s_2, s_3, \dots\}$ (so it is Σ_1 with the element $s_1 = g_1 s$ removed), then $\langle \Sigma_2 \rangle$ is isomorphic to $F(\Sigma_2)$. It follows that the map $\psi(s_n) = s_{n+1}$ gives rise to an isomorphism from $\langle \Sigma_1 \rangle$ to $\langle \Sigma_2 \rangle$. Now take the HNN extension $(G * \mathbb{Z})*_\psi$, whose stable letter we denote by t. This group contains a copy of G, as we noted before. Moreover, since we have ensured that $t s_n t^{-1} = s_{n+1}$ for every $n \geqslant 1$, it can be generated by just the three elements s_1, s, and t. Thus, we have embedded an arbitrary countable group into a group with three generators. (We leave the reader to think about how one can vary this construction to produce a group with two generators.)

5.3 There Are Uncountably Many Nonisomorphic Finitely Generated Groups

This was proved by B. H. Neumann in 1932. Since there are infinitely many primes, there are uncountably many nonisomorphic groups of the form $\bigoplus_{p \in P} \mathbb{Z}_p$, where P is an infinite set of primes. We have seen that each of these groups can be embedded in a finitely generated group, and our earlier comments on finite subgroups of HNN extensions show that no two of the resulting finitely generated groups are isomorphic.

5.4 An Answer to Hopf's Question

A group G is called *Hopfian* if every surjective homomorphism from G to G is an isomorphism. Most familiar groups have this property: for example, finite groups obviously do, as do $\mathbb{Z}^n$ (as you can prove using linear algebra) and free groups. So too do groups of matrices such as $\mathrm{SL}(n, \mathbb{Z})$, as we shall discuss in a moment. An example of a non-Hopfian group is the group of all infinite sequences of integers (under pointwise addition), since the function that takes $(a_1, a_2, a_3, \dots)$ to $(a_2, a_3, a_4, \dots)$ is a surjective homomorphism that contains $(1, 0, 0, \dots)$ in its kernel. But is there a finitely presented example? The answer is yes, and Higman was the first to construct one. The following examples are due to Baumslag and Solitar.

Let $p \geqslant 2$ be an integer and identify $\mathbb{Z}$ with the free group $\langle a \rangle$ generated by a single generator a. Then the subgroups $p\mathbb{Z}$ and $(p+1)\mathbb{Z}$ of $\mathbb{Z}$ are identified with the powers of a^p and a^{p+1}, respectively. Let ψ be the isomorphism between these subgroups that takes a^p to a^{p+1} and consider the corresponding HNN extension B. This has presentation $B = \langle a, t \mid ta^{-p}t^{-1}a^{p+1}\rangle$. The homomorphism $\psi : B \to B$ defined by $t \mapsto t$, $a \mapsto a^p$ is clearly a surjection but its kernel contains, for example, the element $c = ata^{-1}t^{-1}a^{-2}tat^{-1}a$, which does not contain a pinch and is therefore not equal to the identity, by Britton's lemma. (If you want to convince yourself how useful this lemma is, set $p = 3$ and try to prove directly that c is not equal to the identity in the group B just defined.)

5.5 A Group that Has No Faithful Linear Representation

One can show that a finitely generated group G of matrices over any field is *residually finite*, which means that for each nontrivial element $g \in G$ there exists a finite group Q and a homomorphism $\pi : G \to Q$ with $\pi(g) \neq 1$. For example, if you are given an element $g \in \mathrm{SL}(n, \mathbb{Z})$, then you can pick an integer m bigger than the absolute values of all the entries in g (which is an $n \times n$ matrix) and consider the homomorphism from $\mathrm{SL}(n, \mathbb{Z})$ to $\mathrm{SL}(n, \mathbb{Z}/m\mathbb{Z})$ that reduces the matrix entries mod m. The image of g in the finite group $\mathrm{SL}(n, \mathbb{Z}/m\mathbb{Z})$ is clearly nontrivial.

Non-Hopfian groups are not residually finite, and hence are not isomorphic to a group of matrices over any field. One can see that the non-Hopfian group B defined above is not residually finite by considering what happens to the nontrivial element c. We saw that there was a surjective homomorphism $\psi : B \to B$ with $\psi(c) = 1$. Let c_n be an element such that $\psi^n(c_n) = c$ (which exists since ψ is a surjection). If there were a homomorphism π from B to a finite group Q with $\pi(c) \neq 1$, then we would have infinitely many distinct homomorphisms from B to Q, namely the compositions $\pi \circ \psi^n$; these are distinct because $\pi \circ \psi^m(c_n) = 1$ if $m > n$ and $\pi \circ \psi^n(c_n) = \pi(c) \neq 1$. This is a contradiction, since a homomorphism from a finitely generated group to a finite group is determined by what it does to the generators, so there can only be finitely many such homomorphisms.

5.6 Infinite Simple Groups

Britton's lemma actually tells us more than that $c \neq 1$: the subgroup Λ of B generated by t and c is in fact a free group on those generators. Thus we may form the amalgamated free product Γ of two copies of B, denoted B_1 and B_2, by gluing together the two copies of Λ with the isomorphism $c_1 \mapsto t_2$, $t_1 \mapsto c_2$. We have seen that in any finite quotient of $\Gamma = B_1 *_\Lambda B_2$, the elements c_1 $(= t_2)$ and c_2 $(= t_1)$ must have trivial image, and it is easy to deduce from this that in fact the quotient must be trivial. Thus Γ is an infinite group with no finite quotients. It follows that the quotient of Γ by any maximal proper normal subgroup is also infinite (and it is simple by maximality).

The simple group that we have constructed is infinite and finitely generated but it is not finitely presentable. Finitely presented infinite simple groups do exist, but they are much harder to construct.

6 Higman's Theorem and Undecidability

We have seen that there are uncountably many (nonisomorphic) finitely generated groups. But as there are only countably many finitely *presented* groups, only countably many finitely generated groups can be subgroups of finitely presented groups. Which ones are they?

A complete answer to this question is provided by the following beautiful and deep theorem proved by Graham Higman in 1961. It says, roughly, that the groups that arise are all those that are algorithmically describable. (If you have no idea what this means, even roughly, then you might like to read THE INSOLUBILITY OF THE HALTING PROBLEM (**??**) before continuing with this section.)

A set S of words over a finite alphabet A is called *recursively enumerable* if there is some algorithm (or more formally, TURING MACHINE (**??**)) that can produce a complete list of the elements of S. A case of particular interest is when A is just a singleton, in which case a word is determined by its length and we can think of S as a set of nonnegative integers. The elements of S need not be listed in a sensible order, so having an algorithm that produces an exhaustive list of S does not mean that one can use the algorithm to determine that some given word w does *not* belong to S: if you imagine standing by your computer as it enumerates S, there will not in general come a time when you can say to yourself, "If it was going to appear, then it would have done so by now," and therefore be certain that it is not in S. If you want an algorithm with this further property, then you need the stronger notion of a *recursive set*, which is a set S such that S and its complement are *both* recursively enumerable. Then you can list all the elements that belong to S and you can also list all the elements that do not belong to S.

A finitely generated group is said to be *recursively presentable* if it has a presentation with a finite number of generators and a recursively enumerable set of defining relations. In other words, such a group is not necessarily finitely presented, but at least the presentation of the group is "nice" in the sense that it can be generated by some algorithm.

Higman's embedding theorem states that *a finitely generated group G is recursively presentable if and only if it is isomorphic to a subgroup of a finitely presented group.*

To get a feeling for how nonobvious this is, you might consider the following presentation of the group of all rationals under addition, in which the generator a_n corresponds to the fraction $1/n!$:

$$Q = \langle a_1, a_2, \cdots \mid a_n^n = a_{n-1} \ \forall n \geqslant 2 \rangle.$$

Higman's theorem tells us that Q can be embedded in a finitely presented group, but no truly explicit embedding is known.

The power of Higman's theorem is illustrated by the ease with which it implies the celebrated undecidability results that were rightly regarded as watersheds of twentieth-century mathematics. In order to make this case convincingly, I shall give a complete proof (except that I shall assume some of the facts mentioned earlier) that there exist finitely presented groups with unsolvable word problems, and also that there are sequences of finitely presented groups among which one cannot decide isomorphism. We shall also see how these group-theoretic results can be used to translate undecidability phenomena into topology.

The basic seed of undecidability comes from the fact that there are recursively enumerable subsets $S \subset \mathbb{N}$ that are not recursive. Using this fact one can readily construct finitely generated groups with an unsolvable word problem: given such a set of integers S we consider

$$J := \langle a, b, t \mid t(b^n a b^{-n})t^{-1} = b^n a b^{-n} \ \forall n \in S \rangle.$$

This is the HNN extension of the free group $F(a, b)$ associated with the identity map $L \to L$, where L is the subgroup generated by $\{b^n a b^{-n} : n \in S\}$. Britton's lemma tells us that the word $w_m := t(b^m a b^{-m})t^{-1}(b^m a^{-1} b^{-m})$ equals $1 \in J$ if and only if $m \in S$, and by definition there is no algorithm to decide if $m \in S$, so we cannot decide which of the w_m are relations. Thus J has an unsolvable word problem.

That there exist finitely presented groups for which the word problem is unsolvable is a much deeper fact, but with Higman's embedding theorem to hand the proof becomes almost trivial: Higman tells us that J can be embedded in a finitely presented group Γ, and it is a relatively straightforward exercise to show that if one cannot decide which words in the generators of J represent the identity, then one cannot decide for arbitrary words in the generators of Γ either.

Once one has a finitely presented group with an unsolvable word problem, it is easy to translate undecidability into all manner of other problems. For example, suppose that $\Gamma = \langle A \mid R \rangle$ is a finitely presented group with an unsolvable word problem, where $A = \{a_1, \ldots, a_n\}$ and no a_i equals the identity in Γ. For each word w made out of the letters in A and their inverses, define a group Γ_w to have presentation

$$\langle A, s, t \mid R, \ t^{-1}(s^i a_i s^{-i})t(s^i w s^{-i}), \ i = 1, \ldots, n \rangle.$$

It is not hard to show that if $w = 1$ in Γ then Γ_w is the free group generated by s and t. If $w \neq 1$, then Γ_w is an HNN extension. In particular, it contains a copy of Γ, and hence has an unsolvable word problem, which means that it cannot be a free group. Thus, since there is no algorithm to decide whether $w = 1$ in Γ, one cannot decide which of the groups Γ_w are isomorphic to which others.

A variant of this argument shows that there is no algorithm to determine whether or not a given finitely presented group is trivial.

We shall see in a moment that every finitely presented group G is the fundamental group of some compact four-dimensional manifold. By following a standard proof of this theorem with considerable care, Markov proved in 1958 that in dimensions 4 and above there is no algorithm to decide which compact manifolds (presented as simplicial complexes, for example) are homeomorphic. His basic idea was to show that if there were an algorithm to determine which triangulated 4-manifolds are homeomorphic, then one could use it to determine which finitely presented groups are trivial, which we know is impossible. In order to implement this idea one has to be careful to arrange that the 4-manifolds associated with different presentations of the trivial group are homeomorphic: this is the delicate part of the argument.

Strikingly, there does exist an algorithm to decide which compact three-dimensional manifolds are isomorphic. This is an extremely deep theorem that relies in particular on Perelman's solution to THURSTON'S GEOMETRIZATION CONJECTURE (**??**).

7 Topological Group Theory

Let us change perspective now and look at the symbols $P \equiv \langle a_1, \ldots, a_2 \mid r_1, \ldots, r_m \rangle$ through the eyes of a topologist. Instead of interpreting P as a recipe for constructing a group, we regard it as a recipe for constructing a TOPOLOGICAL SPACE (**??**), more specifically a *two-dimensional complex.* Such spaces consist of points, called *vertices*, some of which are linked by directed paths, called *edges*, or 1-*cells.* If a collection of such 1-cells forms a cycle, then it can be filled in with a *face*, or 2-*cell*: topologically speaking, each face is a disk with a directed cycle as its boundary.

To see what this complex is, let us first consider the standard presentation $P \equiv \langle a, b \mid aba^{-1}b^{-1} \rangle$ of $\mathbb{Z}^2$. (This is generated by a and b and the relation tells us that $ab = ba$.) We begin with a graph K^1 that has a single vertex and two edges (which are loops) that are directed and labeled a and b. Next, we take a square $[0, 1] \times [0, 1]$, the sides of which are directed and labeled a, b, a^{-1}, b^{-1} as we proceed around the boundary. Imagine gluing the boundary of the square to the graph so as to respect the labeling of edges: with a bit of thought, you should be able to see that the result is a torus, that is, a surface in the shape of a bagel. An observation that turns out to be important is that the fundamental group of the torus is $\mathbb{Z}^2$, the group we started with.

The idea of "gluing" is made precise by the use of *attaching maps*: we take a continuous map ϕ from the boundary of the square S to the graph K^1 that sends the corners of the square to the vertex of K^1 and sends each side (minus its vertices) homeomorphically onto an open edge. The torus is then the quotient of $K^1 \sqcup S$ by the equivalence relation that identifies each x in the boundary of the square with its image $\phi(x)$.

With this more abstract language in hand, it is easy to see how the above construction generalizes to arbitrary presentations: given a presentation $P \equiv \langle a_1, \dots, a_n \mid r_1, \dots, r_m \rangle$, one takes a graph with a single vertex and n oriented loops, which are labeled $a_1, \dots, a_n$. Then for each r_j one attaches a polygonal disk by gluing its boundary circuit to the sequence of oriented edges that traces out the word r_j.

In general, the result will not be a surface as it was for $\langle a, b \mid aba^{-1}b^{-1} \rangle$. Rather, it will be a two-dimensional complex with singularities along the edges and at the vertex. You may find it instructive to do some more examples. From $\langle a \mid a^2 \rangle$ one gets the projective plane; from $\langle a, b, c, d \mid aba^{-1}b^{-1}, cdc^{-1}d \rangle$ one gets a torus and a Klein bottle stuck together at a point. Picturing the 2-complex for $\langle a, b \mid a^2, b^3, (ab)^3 \rangle$ is already rather difficult.

The construction of $K(P)$ is the beginning of *topological group theory*. The Seifert–van Kampen theorem (mentioned earlier) implies that the fundamental group of $K(P)$ is the group presented by P. But the group no longer sits inertly in the form of an inscrutable presentation—now it acts on the UNIVERSAL COVERING (**??**) of $K(P)$ by homeomorphisms known as "deck transformations." Thus, through the simple construction of $K(P)$ (and the elegant theory of covering spaces in topology) we achieve our aim of realizing an abstract finitely presented group as the group of symmetries of an object with a potentially rich structure, on which we can bring global geometric and topological techniques to bear.

To obtain an improved topological model for our group, we can embed $K(P)$ in $\mathbb{R}^5$ (just as one can embed a finite GRAPH (**??**) in $\mathbb{R}^3$) and consider the compact four-dimensional manifold M obtained by taking all points that are a small fixed distance from the image. (I am assuming that the embedding is suitably "tame," which one can arrange.) The mental picture to strive for here is a higher-dimensional analogue of the surface (sleeve) one gets by taking the points in $\mathbb{R}^3$ a small fixed distance from an embedded graph. The fundamental group of M is again the group presented by P, so now we have our arbitrary finitely presented group acting on a manifold (the universal cover of M). This allows us to use the tools of analysis and DIFFERENTIAL GEOMETRY (**??**).

The constructions of $K(P)$ and M establish the more difficult implication of the theorem, promised earlier, that a group can be finitely presented if and only if it is the fundamental group of a compact cell complex and of a compact 4-manifold. This result raises several natural questions. First, are there better, more informative, topological models for an arbitrary finitely presented group Γ? And if not, then what can one say about the classes of groups defined by the natural constraints that arise when one tries to improve the model? For example, we would like to construct a lower-dimensional manifold with fundamental group Γ, enabling us to exploit our physical insight into three-dimensional geometry. But it turns out that the fundamental groups of compact three-dimensional manifolds are very special; this observation lies near the heart of a great deal of mathematics at the end of the twentieth century. Other interesting fields open up when one asks which groups arise as the fundamental groups of compact spaces satisfying CURVATURE (**??**) conditions, or constraints coming from complex geometry.

A particularly rich set of constraints comes from the following question. Can one arrange for an arbitrary finitely presented group to be the fundamental group of a compact space (a complex or manifold, perhaps) whose universal cover is CONTRACTIBLE (**??**)? This is a natural question from the point of view of topology because a space with a contractible universal cover is, up to HOMOTOPY (**??**), completely determined by its fundamental group. If the fundamental group is Γ, then such a space is called a *classifying space* for Γ and its homotopy-invariant properties provide a rich array of invariants for the group Γ (getting away from the gross dependence that $K(P)$ has on P rather than Γ).

If our earlier discussion of how hard it is to recognize Γ from P has left you very skeptical about whether this dependence can actually be removed, then your skepticism is well-founded: there are many obstructions to the construction of compact classifying spaces for an arbitrary finitely presented group; the study of them (under the generic name *finiteness conditions*) is a rich area at the interface of modern group theory, topology, and homological algebra.

One aspect of this area is the search for natural conditions that ensure the *existence* of compact classifying spaces (not necessarily manifolds). This is one of several places where manifestations of nonpositive curvature play a fundamental role in modern group theory. More combinatorial conditions also arise. For example, Lyndon proved that for any presentation $P \equiv \langle A \mid r\rangle$ where the single defining relation $r \in F(A)$ is not a nontrivial power, the universal cover of $K(P)$ is contractible.

A neighboring and highly active area of research concerns questions of uniqueness and rigidity for classifying spaces. (Here, as is common, the word *rigidity* is used to describe a situation in which requiring two objects to be equivalent in an apparently weak sense forces them to be equivalent in an apparently stronger sense.) For example, the (open) *Borel conjecture* asserts that if two compact manifolds have isomorphic fundamental groups and contractible universal covers, then those manifolds must be homeomorphic.

I have been talking mostly about realizing groups as fundamental groups, which led to certain free actions. That is, we could interpret the elements of the group as symmetries of a topological space and none of these symmetries had any fixed points. Before moving on to geometric group theory I should point out that there are many situations in which the most illuminating actions of a group are not free: one instead allows well-understood stabilizers. (The *stabilizer* of a point is the set of all symmetries in the group that leave that point fixed.) For example, the natural way in which to study Γ_Δ is by its action on the triangulated plane, each vertex of which is left unmoved by twelve symmetries.

A deeper illustration of the merits of seeking insight into algebraic structure through nonfree actions on suitable topological spaces comes from the Bass–Serre theory of groups acting on trees, which subsumes the theory of amalgamated free products and HNN extensions, whose potency we saw earlier. (This theory and its extensions often go under the heading of *arboreal group theory.*)

A *tree* is a connected graph that has no circuits in it. It is helpful to regard it as a metric space in which each edge has length 1. The group actions that one allows on trees are those that take edges to edges isometrically, never flipping an edge.

If a group Γ acts on a set X (in other words, if it can be regarded as a group of symmetries of X), then the *orbit* of a point $x \in X$ is the set of all its images gx with $g \in \Gamma$. A group Γ can be expressed as an amalgamated free product $A*_C B$ if and only if it acts on a tree in such a way that there are two orbits of vertices, one orbit of edges, and stabilizers A, B, C (where A and B are the stabilizers of adjacent vertices and intersect in C, which is the edge stabilizer). HNN extensions correspond to actions with one orbit of vertices and one orbit of edges. Thus, amalgamated free products and HNN extensions appear as *graphs of groups*, which are the basic objects of Bass–Serre theory. These objects allow one to recover groups acting on trees from the quotient data of the action, i.e., the quotient space (which is a graph) and the pattern of edge and vertex stabilizers.

An early benefit of Bass–Serre theory is a transparent and instructive proof that any finite subgroup of $A *_C B$ is conjugate to a subgroup of either A or B: given any set V of vertices in a tree, there is a unique vertex or midpoint x minimizing $\max\{d(x,v) \mid v \in V\}$; one applies this observation with V an orbit of the finite subgroup; x provides a fixed point for the action of the subgroup; and any point stabilizer is conjugate to a subgroup of either A or B.

Arboreal group theory goes much deeper than this first application suggests. It is the basis for a decomposition theory of finitely presented groups from which it emerges, for example, that there is an essentially canonical maximal splitting of an arbitrary finitely presented group as a graph of groups with cyclic edge stabilizers. This provides a striking parallel with the decomposition theory of 3-manifolds, a parallel that extends far beyond a mere analogy and accounts for much of the deepest work in geometric group theory in the past ten years. If you want to learn more about this, search the literature for *JSJ decompositions*. You may also want to search for *complexes of groups*, which provide the appropriate higher-dimensional analogue for graphs of groups.

8 Geometric Group Theory

Let us refresh the image of $K(P)$ in our mind's eye by thinking again about the presentation $P \equiv \langle a,b \mid aba^{-1}b^{-1}\rangle$ of $\mathbb{Z}$. The complex $K(P)$, as we saw earlier, is a torus. Now the torus can be defined as the quotient of the Euclidean plane $\mathbb{R}^2$ by the action of the group $\mathbb{Z}^2$ (where the point $(m,n) \in \mathbb{Z}^2$ acts as the translation $(x,y) \mapsto (x+m, y+n)$): in fact, $\mathbb{R}^2$, with an appropriate square tiling, is the UNIVERSAL COVER (??) of the torus. If we look at the orbit of the point 0 under this action, it forms a copy of $\mathbb{Z}^2$, and one can thereby see the large-scale geometry of $\mathbb{Z}^2$ laid out for us. We can make the idea of the "geometry of $\mathbb{Z}^2$" precise by decreeing that edges of the tiling have length 1 and defining the *graph distance* between vertices to be the length of the shortest path of edges connecting them.

As this example shows, the construction of $K(P)$ involves the two main (intertwined) strands of geometric group theory. In the first and more classical strand, one studies actions of groups on metric and topological spaces in order to elucidate the structures of both the space and the group (as with the action of $\mathbb{Z}^2$ on the plane in our example, or the action of the fundamental group of $K(P)$ on its universal cover in general). The quality of the insights that one obtains varies according to whether the action has or does not have certain desirable properties. The action of $\mathbb{Z}^2$ on $\mathbb{R}^2$ consists of isometries on a space with a fine geometric structure, and the quotient (the torus) is compact. Such actions are in many ways ideal, but sometimes one accepts weaker admission criteria in order to obtain a more diverse class of groups, and sometimes one demands even more structure in order to narrow the focus and study groups and spaces of an exceptional, but for that reason interesting, character.

This first strand of geometric group theory mingles with the second. In the second strand, one regards finitely generated groups as geometric objects in their own right equipped with *word metrics*, which are defined as follows. Given a finite generating set S for a group Γ, one defines the *Cayley graph* of Γ by joining each element $\gamma \in \Gamma$ by an edge to each element of the form γs or γs^{-1} with $s \in S$ (which is the same as the graph formed by the edges of the universal covering of $K(P)$). The distance $d_S(\gamma_1, \gamma_2)$ between γ_1 and γ_2 is then the length of the shortest path from γ_1 to γ_2 if all edges have length 1. Equivalently, it is the length of the shortest word in the free group on S that is equal to $\gamma_1^{-1}\gamma_2$ in Γ.

The word metric and Cayley graph depend on the choice of generating set but their large-scale geometry do not. In order to make this idea precise, we introduce the notion of a *quasi-isometry*. This is an equivalence relation that identifies spaces that are similar on a large scale. If X and Y are two metric spaces, then a quasi-isometry from X to Y is a function $\phi : X \to Y$ with the following two properties. First, there are positive constants c, C, and ϵ such that $cd(x, y) - \epsilon \leqslant d(\phi(x), \phi(y)) \leqslant Cd(x, y) + \epsilon$: this says that ϕ distorts sufficiently large distances by at most a constant factor. Second, there is a constant C' such that for every $y \in Y$ there is some $x \in X$ for which $d(\phi(x), y) \leqslant C'$: this says that ϕ is a "quasi-surjection" in the sense that every element of Y is close to the image of an element of X.

Consider for example the two spaces $\mathbb{R}^2$ and $\mathbb{Z}^2$, where the metric on $\mathbb{Z}^2$ is given by the graph distance defined earlier. In this case the map $\phi : \mathbb{R}^2 \to \mathbb{Z}^2$ that takes (x, y) to $(\lfloor x \rfloor, \lfloor y \rfloor)$ (where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x) is easily seen to be a quasi-isometry: if the Euclidean distance d between two points (x, y) and (x', y') is at least 10, say, then the graph distance between $(\lfloor x \rfloor, \lfloor y \rfloor)$ and $(\lfloor x' \rfloor, \lfloor y' \rfloor)$ will certainly lie between $\frac{1}{2}d$ and $2d$. Notice how little we care about the local structure of the two spaces: the map ϕ is a quasi-isometry despite not even being continuous.

It is not hard to check that if ϕ is a quasi-isometry from X to Y, then there is a quasi-isometry ψ from Y to X that "quasi-inverts" ϕ, in the sense that every x in X is at most a bounded distance from $\psi\phi(x)$ and every y in Y is at most a bounded distance from $\phi\psi(y)$. Once one has established this, it is easy to see that quasi-isometry is an equivalence relation.

Returning to Cayley graphs and word metrics, it turns out that if you take two different sets of generators for the same group, then the resulting Cayley graphs will be quasi-isometric. Thus, any property of a Cayley graph that is invariant under quasi-isometry will be a property not just of the graph but of the group itself. When dealing with such invariants we are free to think of Γ itself as a space (since we do not care which Cayley graph we form), and we can replace it by any metric space that is quasi-isometric to it, such as the universal cover of a closed Riemannian manifold with fundamental group Γ (whose existence we discussed earlier). Then the tools of analysis can be brought to bear.

A fundamental fact, discovered independently by many people and often called the *Milnor–Švarc lemma*, provides a crucial link between the two main strands of geometric group theory. Let us call a metric space X a *length space* if the distance between each pair of points is the infimum of the lengths of paths joining them. The Milnor–Švarc lemma states that if a group Γ acts nontrivially as a set of isometries of a length space X, and if the quotient is compact, then Γ is finitely generated and quasi-isometric to X (for any choice of word metric).

We have seen an example of this already: $\mathbb{Z}^2$ is quasi-isometric to the Euclidean plane. Less obviously, the same is true of Γ_Δ. (Consider the map that takes each element α of Γ_Δ to the point of $\mathbb{Z}^2$ nearest $\alpha(0)$.)

The fundamental group of a compact Riemannian manifold is quasi-isometric to the universal cover of that manifold. Therefore, from the point of view of quasi-isometry invariants, the study of such manifolds is equivalent to the study of arbitrary finitely presented groups. In a moment we will discuss some nontrivial consequences of this equivalence. But first let us reflect on the fact that, when finitely generated groups are considered as metric objects in the framework of large-scale geometry, they present us with a new challenge: we should *classify finitely generated groups up to quasi-isometry*.

This is an impossible task, of course, but nevertheless serves as a guiding beacon in modern geometric group theory—a beacon that has guided us towards many beautiful theorems, particularly under the general heading of rigidity. For example, suppose that you come across a finitely generated group Γ that is reminiscent of $\mathbb{Z}^n$ on a large scale: in other words, quasi-isometric to it. We are not necessarily given any algebraically defined map between this mystery group and $\mathbb{Z}^n$, and yet it transpires that such a group must contain a copy of $\mathbb{Z}^n$ as a subgroup of finite index.

At the heart of this result is *Gromov's polynomial growth theorem*, a landmark theorem published in 1981. This theorem concerns the number of points within a distance r of the identity in a finitely generated group Γ. This will be a function $f(r)$, and Gromov was interested in how the function $f(r)$ grows as r tends to infinity, and what that tells us about the group Γ.

If Γ is an Abelian group with d generators, then it is not hard to see that $f(r)$ is at most $(2r+1)^d$ (since each generator is raised to a power between $-r$ and r). Thus, in this case $f(r)$ is bounded above by a polynomial in r. At the other extreme, if Γ is a free group with two generators a and b, say, then $f(r)$ is exponentially large, since all sequences of length r that consist of as and bs (and not their inverses) give different elements of Γ.

Given this sharp contrast in behavior, one might wonder whether requiring $f(r)$ to be bounded above by a polynomial forces Γ to exhibit a great deal of commutativity. Fortunately, there is a much-studied definition that makes this idea precise. Given any group G and any subgroup H of G, the *commutator* $[G,H]$ is the subgroup generated by all elements of the form $ghg^{-1}h^{-1}$, where g belongs to G and h belongs to H. If G is Abelian, then $[G,H]$ contains just the identity. If G is not Abelian, then $[G,G]$ forms a group G_1 that contains other elements besides the identity, but it may be that $[G,G_1]$ is trivial. In that case, one says that G is a two-step nilpotent group. In general, a *k-step nilpotent* group G is one where, if you form a sequence by setting $G_0 = G$ and $G_{i+1} = [G,G_i]$ for each i, then you eventually reach the trivial group, and the first time you do so is at G_k. A *nilpotent* group is a group that is k-step nilpotent for some k.

Gromov's theorem states that a group has polynomial growth if and only if it has a nilpotent subgroup of finite index. This is a quite extraordinary fact: the polynomial growth condition is easily seen to be independent of the choice of word metric and to be an invariant of quasi-isometry. Thus the seemingly rigid and purely algebraic condition of having a nilpotent subgroup of finite index is in fact a quasi-isometry invariant, and therefore a flabby, robust characteristic of the group.

In the past fifteen years quasi-isometric rigidity theorems have been established for many other classes of groups, including lattices in semisimple Lie groups and the fundamental groups of compact 3-manifolds (where the classification up to quasi-isometry involves more than algebraic equivalences), as well as various classes defined in terms of their graph of group decompositions. In order to prove theorems of this type, one must identify nontrivial invariants of quasi-isometry that allow one to distinguish and relate various classes of spaces. In many cases such invariants come from the development of suitable analogues of the tools of algebraic topology, modified so that they behave well with respect to quasi-isometries rather than continuous maps.

9 The Geometry of the Word Problem

It is time to explain the comments I made earlier about the geometry inherent in the basic decision problems of combinatorial group theory. I shall concentrate exclusively on the geometry of the word problem.

Gromov's *filling theorem* describes a startlingly intimate connection between the highly geometric study of disks with minimal area in RIEMANNIAN GEOMETRY (**??**) and the study of word problems, which seems to belong more to algebra and logic.

On the geometric side, the basic object of study is the *isoperimetric function* $\mathrm{Fill}_M(l)$ of a smooth compact manifold M. Given any closed path of length l, there is a disk of minimal area that is bounded by that path. The largest such area, over all closed paths of length l, is defined to be $\mathrm{Fill}_M(l)$. Thus, the isoperimetric function is the smallest function of which it is true to say that every closed path of length l can be filled by a disk of area at most $\mathrm{Fill}_M(l)$.

The image to have in mind here is that of a soap film: if one twists a circular wire of length l in Euclidean space and dips it in soap, the film that forms has area at most $l^2/4\pi$, whereas if one performs the same experiment in HYPERBOLIC SPACE (**??**), the area of the film is bounded by a linear function of l. Correspondingly, the isoperimetric functions of $\mathbb{E}^n$ and $\mathbb{H}^n$ (and quotients of them by groups of isometries) are quadratic and linear, respectively. In a moment we shall discuss what types of isoperimetric functions arise when one considers other geometries (more precisely, compact Riemannian manifolds).

To state the filling theorem we need to think about the algebraic side as well. Here, we identify a function that measures the complexity of a direct attack on the word problem for an arbitrary finitely presented group $\Gamma = \langle A \mid R\rangle$. If we wish to know whether a word w equals the identity in Γ and do not have any further insight into the nature of Γ, then there is not much we can do other than repeatedly insert or remove the given relations $r \in R$.

Consider the simple example $\Gamma = \langle a, b \mid b^2a, baba\rangle$. In this group aba^2b represents the identity. How do we prove this? Well,

$$\begin{aligned} aba^2b &= a(b^2a)ba^2b = ab(baba)ab \\ &= abab = a(baba)a^{-1} = aa^{-1} = 1. \end{aligned}$$

Now let us think about the proof geometrically, via the Cayley graph. Since $aba^2b = 1$ in the group Γ, we obtain a cycle in this graph if we start at the identity and go along edges labeled a, b, a, a, b, in that order (in which case we visit the vertices 1, a, ab, aba, aba^2, $aba^2b = 1$). The equalities in the proof can be thought of as a way of "contracting" this cycle down to the identity by means of inserting or deleting small loops: for instance, we could insert b, a, b, a into the list of edge directions, since $baba$ is a relation, or we could delete a trivial loop of the form a, a^{-1}. This contraction can be given a more topological character if we turn our Cayley graph into a two-dimensional complex by filling in each small loop with a *face*. Then the contraction of the original cycle consists in gradually moving it across these small faces.

Thus, the difficulty of demonstrating that a word w equals the identity is intimately connected with the *area* of w, denoted $\text{Area}(w)$, which can be thought of algebraically as the smallest sequence of relations you need to insert and delete to turn w into the identity, or geometrically as the smallest number of faces you need to make a disk that fills the cycle represented by w.

The *Dehn function* $\delta_\Gamma : \mathbb{N} \to \mathbb{N}$ bounds $\text{Area}(w)$ in terms of the length $|w|$ of the word w: $\delta_\Gamma(n)$ is the largest area of any word of length at most n that equals 1 in Γ. If the Dehn function grows rapidly, then the word problem is hard, since there are short words that are equal to the identity, but their area is very large, so that any demonstration that they are equal to the identity has to be very long. Results bounding the Dehn function are called *isoperimetric inequalities*.

The subscript on δ_Γ is somewhat misleading since different finite presentations of the same group will in general yield different Dehn functions. This ambiguity is tolerated because it is tightly controlled: if the groups defined by two finite presentations are isomorphic, or just quasi-isometric, then the corresponding Dehn functions have similar growth rates. More precisely, they are *equivalent*, with respect to what is sometimes called the *standard equivalence relation* "$\simeq$" of geometric group theory: given two monotone functions $f, g : [0,\infty) \to [0,\infty)$, one writes $f \preccurlyeq g$ if there exists a constant $C > 0$ such that $f(l) \leqslant Cg(Cl+C)+Cl+C$ for all $l \geqslant 0$, and $f \simeq g$ if $f \preccurlyeq g$ and $g \preccurlyeq f$; and one extends this relation to include functions $\mathbb{N} \to [0,\infty)$.

You will have noticed a resemblance between the definitions of $\text{Fill}_M(l)$ and $\delta_\Gamma(n)$. The filling theorem relates them precisely: it states that *if M is a smooth compact manifold, then $\text{Fill}_M(l) \simeq \delta_\Gamma(l)$, where Γ is the fundamental group $\pi_1 M$ of M.*

For example, since $\mathbb{Z}^2$ is the fundamental group of the torus $T = \mathbb{R}^2/\mathbb{Z}^2$, which has Euclidean geometry, $\delta_{\mathbb{Z}^2}(l)$ is quadratic.

9.1 What Are the Dehn Functions?

We have seen that the complexity of word problems is related to the study of isoperimetric problems in Riemannian and combinatorial geometry. Such insights have, in the last fifteen years, led to great advances in the understanding of the nature of Dehn functions. For example, one can ask for which numbers ρ the function n^ρ is a Dehn function. The set of all such numbers, which can be shown to be countable, is known as the *isoperimetric spectrum*, denoted IP, and it is now largely understood.

Following work by many authors, Brady and Bridson proved that the closure of IP is $\{1\} \cup [2,\infty)$. The finer structure of IP was described by Birget, Rips, and Sapir in terms of the time functions of Turing machines. A further result by the same authors and Ol'shanskii explains how fundamental Dehn functions are to understanding the complexity of arbitrary approaches to the word problem for finitely generated groups Γ: the word problem for Γ lies in NP if and only if Γ is a subgroup of a finitely presented group with polynomial Dehn function. (Here, NP is the class of problems in the famous "P = NP" question: see COMPUTATIONAL COMPLEXITY (**??**) for a description of this class.)

The structure of IP raises an obvious question: What can one say about the two classes of groups singled out as special—those with linear Dehn functions and those with quadratic ones? The true nature of the class of groups with a quadratic Dehn function remains obscure for the moment but there is a beautifully definitive description of those with a linear Dehn function: they are the *word hyperbolic groups*, which we shall discuss in the next section.

Not all Dehn functions are of the form n^α: there are Dehn functions such as $n^\alpha \log n$, for example, and others that grow more quickly than any iterated exponential, for example that of

$$\langle a, b \mid aba^{-1}bab^{-1}a^{-1}b^{-2} \rangle.$$

If Γ has unsolvable word problem, then $\delta_\Gamma(n)$ will grow faster than any recursive function (indeed this serves as a definition of such groups).

9.2 The Word Problem and Geodesics

A *closed geodesic* on a Riemannian manifold is a loop that locally minimizes distance, such as a loop formed by an elastic band when released on a perfectly smooth surface. Examples such as the great circles on a sphere or the waist of an hourglass show that manifolds may contain closed geodesics that are *null-homotopic*: that is, they can be moved continuously until they are reduced to a point. But can one construct a compact topological manifold with the property that no matter what metric one puts on it there will always be infinitely many such geodesics? (Technically, if you go round a geodesic loop n times, then you get a geodesic; we avoid this by counting only "primitive" geodesics.)

From a purely geometric point of view this is a daunting problem: all specific metric information has been stripped away and one has to deal with an arbitrary metric on the floppy topological object left behind. But group theory provides a solution: *if the Dehn function of the fundamental group* $\pi_1 M$ *grows at least as fast as* 2^{2^n}*, then in any Riemannian metric on* M *there will be infinitely many closed geodesics that are null-homotopic*. The proof of this is too technical to sketch here.

10 Which Groups Should One Study?

Several special classes of groups have emerged from our previous discussion, such as nilpotent groups, 3-manifold groups, groups with linear Dehn functions, and groups with a single defining relation. Now we shall change viewpoint and ask which groups present themselves for study as we set out to explore the universe of all finitely presented groups, starting with the easiest ones.

The trivial group comes first, of course, followed by the finite groups. Finite groups are discussed in various other places in this volume, so I shall ignore them in what follows and adopt the approach of large-scale geometry, blurring the distinction between groups that have a common subgroup of finite index.

The first infinite group is surely $\mathbb{Z}$, but what comes next is open to debate. If one wants to retain the safety of commutativity, then finitely generated Abelian groups come next. Then, as one slowly relinquishes commutativity and control over growth and constructibility, one passes through the progressively larger classes of nilpotent, polycyclic, solvable, and elementary amenable groups. We have already met nilpotent groups in our discussion of Gromov's polynomial-growth theorem. They crop up in many contexts as the most natural generalization of Abelian groups and much is known about them, not least because one can prove a great deal by induction on the k for which they are k-step nilpotent. One can also exploit the fact that G is built from the finitely generated Abelian groups G_i/G_{i+1} in a very controlled way. The larger class of polycyclic groups are built in a similar way, while finitely generated solvable groups are built in a finite number of steps from Abelian groups that need not be finitely generated. This last class is not only larger but wilder; the isomorphism problem is solvable among polycyclic groups, for example, but unsolvable among solvable groups. By definition a group G is solvable if its *derived series*, defined inductively by $G^{(n)} = [G^{(n-1)}, G^{(n-1)}]$ with $G^{(0)} = G$, terminates in a finite number of steps.

The concept known as *amenability* forms an important link between geometry, analysis, and group theory. Solvable groups are amenable but not vice versa. It is not quite the case that a finitely presented group is amenable if and only if it does not contain a free subgroup of rank 2, but for a novice this serves as a good rule of thumb.

Now, let us return to $\mathbb{Z}$ in a more adventurous frame of mind, throw away the security of commutativity, and start taking free products instead. In this more liberated approach, finitely generated free groups appear after $\mathbb{Z}$ as the first groups in the universe. What comes next? Thinking geometrically, we might note that free groups are precisely those groups that have a tree as a Cayley graph and then ask which groups have Cayley graphs that are *tree-like*.

A key property of a tree is that all of its triangles are degenerate: if you take any three points in the tree and join them by shortest paths, then every point in one of these paths is contained in at least one other path as well. This is a manifestation of the fact that trees are spaces of infinite negative curvature. To get a feeling for why, consider what happens when one rescales the metric on a space of bounded negative curvature such as the hyperbolic plane $\mathbb{H}^2$. If we replace the standard distance function $d(x, y)$ by $(1/n)d(x, y)$ and let n tend to ∞, then the curvature of this space (in the classical sense of differential geometry) tends to $-\infty$. This is captured by the fact that triangles look increasingly degenerate: there is a constant $\delta(n)$, with $\delta(n) \to 0$ as $n \to \infty$, such that any side of a triangle in the scaled hyperbolic space $(\mathbb{H}^2, (1/n)d)$ is contained in the $\delta(n)$-neighborhood of the union of the other two sides. More colloquially, triangles in $\mathbb{H}^2$ are *uniformly thin* and get increasingly thin as one rescales the metric.

With this picture in mind, one might move a little away from trees by asking which groups have Cayley graphs in which all triangles are uniformly thin. (It makes little sense to specify the thinness constant δ since it will change when one changes generating set.) The answer is Gromov's *hyperbolic groups*. This is a fascinating class of groups that has many equivalent definitions and arises in many contexts. For example, we already met it as the class of groups that have linear Dehn functions. (It is not at all obvious that these two definitions are equivalent.)

Gromov's great insight is that because the thin-triangles condition encapsulates so much of the essence of the large-scale geometry of negatively curved manifolds, hyperbolic groups share many of the rich properties enjoyed by the groups that act nicely by isometries on such spaces. Thus, for example, hyperbolic groups have only finitely many conjugacy classes of finite subgroups, contain no copy of $\mathbb{Z}^2$, and (after accounting for torsion) have compact classifying spaces. Their conjugacy problems can be solved in less than quadratic time, and Sela showed that one can even solve the isomorphism problem among torsion-free hyperbolic groups. In addition to their many fascinating properties and natural definition, a further source of interest in hyperbolic groups is the fact that in a precise statistical sense, a *random finitely presented group* will be hyperbolic.

Spaces of negative and nonpositive curvature have played a central role in many branches of mathematics in the last twenty years. There is no room even to begin to justify this assertion here but it does guide us in where to look for natural enlargements of the class of hyperbolic groups: we want *nonpositively curved groups*, defined by requiring that their Cayley graphs enjoy a key geometric feature that cocompact groups of isometries inherit from simply connected spaces of nonpositive curvature ("CAT(0) spaces"). But in contrast to the hyperbolic case, the class of groups that one obtains varies considerably when one perturbs the definition, and delineating the resulting classes and their (rich) properties has been the subject of much research.

The added complications that one encounters when one moves from negative to nonpositive curvature are exemplified by the fact that the isomorphism problem is unsolvable in one of the most prominent classes that arises: the so-called *combable groups*.

Let us now return to free groups and ask which hyperbolic groups are the *immediate* neighbors of free groups. Remarkably, this vague question has a convincing answer.

One of the great triumphs of arboreal group theory is the proof that there is a finite description of the set $\mathrm{Hom}(G, F)$ of homomorphisms from an arbitrary finitely generated group G to a free group F. The basic building blocks in this description are what Sela calls *limit groups*. One of the many ways of defining a limit group L is that for each finite subset $X \subset L$ there should exist a homomorphism to a finitely generated free group that is injective on X.

Limit groups can also be defined as those whose first-order logic (see MODEL THEORY (**?? ??**)) resembles that of a free group in a precise sense. To see how first-order logic can be used to say something nontrivial about a group, consider the sentence

$$\forall x, y, z \quad (xy \neq yx) \vee (yz \neq zy) \vee (xz = zx) \vee (y = 1).$$

A group with this property is *commutative transitive*: if x commutes with $y \neq 1$, and y commutes with z, then x commutes with z. Free groups and Abelian groups have this property but a direct product of non-Abelian free groups, for example, does not.

It is a simple exercise to show that free Abelian groups are limit groups. But if one restricts attention to groups that have precisely the same first-order logic as free groups, one gets a smaller class consisting only of hyperbolic groups. The groups in this class are the subject of intense scrutiny at the moment. They all have negatively curved two-dimensional classifying spaces, built from graphs and hyperbolic surfaces in a hierarchical manner. The fundamental groups Σ_g of closed surfaces of genus $g \geqslant 2$ lie in this class, lending substance to the traditional opinion in combinatorial group theory that, among nonfree groups, it is the groups Σ_g that resemble free groups F_n most closely.

Incorporating this opinion into our earlier discussion, we arrive at the view that the groups $\mathbb{Z}^n$, the free groups F_n, and the groups Σ_g are the most basic of infinite groups. This is the start of a rich vein of ideas involving the automorphisms of these groups. In particular, there are many striking parallels between their outer automorphism groups $\mathrm{GL}(n, \mathbb{Z})$, $\mathrm{Out}(F_n)$, and $\mathrm{Mod}_g \cong \mathrm{Out}(\Sigma_g)$ (the mapping class group). These three classes of groups play a fundamental role across a broad spectrum of mathematics. I have mentioned them here in order to make the point that, beyond the search for knowledge about natural classes of groups, there are certain "gems" in group theory that merit a deep and penetrating study in their own right. Other groups that people might suggest for this category include COXETER GROUPS (**??**) (generalized reflection groups, for which Γ_Δ is a prototype) and Artin groups (particularly BRAID GROUPS (**??**), which again crop up in many branches of mathematics).

I have thrown classes of groups at you thick and fast in this last section. Even so, there are many fascinating classes of groups and important issues that I have ignored completely. But so it must be, for as Higman's theorem assures us, the challenges, joys, and frustrations of finitely presented groups can never be exhausted.

Further Reading

Bridson, M. R., and A. Haefliger. 1999. *Metric Spaces of Non-Positive Curvature*. Grundlehren der Mathematischen Wissenschaften, volume 319. Berlin: Springer.

Gromov, M. 1984. Infinite groups as geometric objects. In *Proceedings of the International Congress of Mathematicians, Warszawa, Poland, 1983*, volume 1, pp. 385–92. Warsaw: PWN.

——. 1993. Asymptotic invariants of infinite groups. In *Geometric Group Theory*, volume 2. London Mathematical Society Lecture Note Series, volume 182. Cambridge: Cambridge University Press.

Lyndon, R. C., and P. E. Schupp. 2001. *Combinatorial Group Theory*. Classics in Mathematics. Berlin: Springer.