University of Tasmania
School of Mathematics and Physics

Paradoxes, Monsters, and the Edge of Reason

Joshua T. Deprez, BComp-BSc

November 6, 2009

Supervisor: Dr Des Fitzgerald

Submitted in partial fulfilment of the requirements for the Degree of
Bachelor of Science with Honours

Abstract

A highly surprising result of theoretical mathematics from the last century has been the Banach-Tarski paradox. Starting by finding a rank-two free group among the subgroups of the rotation group SO_3 in ordinary three-dimensional space $\mathbb{R}^3$, a non-intuitive madness results: a sphere can be split into subsets and reassembled using only rotations into *two* copies of the original. Furthermore, almost any two bounded subsets of $\mathbb{R}^3$ may be mutually equidecomposed under the action of rotations and translations. This startling result, with proof, is presented in a modestly accessible form with some illustrations.

A conjecture posited shortly after the publication of the paradox held that only groups with a free subgroup of at least rank 2 could induce a complete lack of group-action invariant finitely-additive measures with finite total measure. A counterexample, the first Tarski monster, was formulated only as recently as 1980 by Ol'shanskii using geometric group theoretic techniques. As a basis for understanding this work, fundamental groups of surfaces and graphs are presented, as well as van Kempen's lemma and diagrams over groups, leading into the small cancellation theory underpinning Tarski monsters. Finally, some topics are discussed relating to the geometry and self-similarity of free groups, with comparison to L-systems, including hyperbolicity and growth conditions.

Acknowledgements

Big thanks first to Des, not least for putting up with my confused emails.

Secondly, big thanks to the School, the staff and postgrad kitchen on the third floor, in particular the electric frypan and water cooler, was truly useful.

A hefty thanks to the Scholarships Office, it came in handy more than once over these past five years.

Thanks to the creators of the open-source Inkscape program, without you my diagrams would be a dreary mess. Thanks also to the creators of LaTeX.

Finally, a huge gushing thanks to my friends Alex Berry and Chris Neugebauer, your comments have really helped.

Contents

List of Figures

Chapter 1

Introduction

It is a truth universally acknowledged that a thesis in possession of some content must be in want of an introduction.[1]

Many things are not so clear. Consider your favourite paradox as an example. A paradox has two faces: an opinion held as intuitively reasonable, and some superior logical argument as to the fallacy of such an opinion. Intuition doesn't give in easily, but once the argument has been understood, intuition is keener to adjust accordingly, and thus the first benefit to understanding and accepting a paradox is to overcome one's often-misleading intuition. The crazier the paradox, the better. How much more crazy than the duplication of balls via isometric transformations can one get?

An initial goal of this work was to comprehend and re-present the Banach-Tarski paradox, including a proof. This goal was not particularly original (Wagon 1985). A distinctive problem with explaining the paradox is the lack of instructive illustrations. The subsets into which a sphere must be separated are difficult to describe (and are shown to exist in a non-constructive manner). Where possible, another aim was to provide illustrations of the processes involved in the proof.

A dominating feature of the proof of the paradox is the way in which a free group of rank 2 is used. The free group has a number of interesting properties that relate the paradox to modern and recent work. This is the second benefit to understanding a paradox—gaining additional tricks to use elsewhere, and noting parallels between the reasoning involved here and

[1]For further reading on a different kind of monster, see the marvellous novel by Austen and Grahame-Smith (2009).

other places.

The Banach-Tarski paradox makes a statement that lies ostentibly inside geometry, but starts with some manipulation of somewhat arbitrary symbols. Like many great proofs, it embeds the theory of one thing within the theory of another. The second aim, then, is to find out what other interesting ways are there to link the abstract symbol-shuffling facet of group theory to the geometric facet. There are numerous. A few are related to elements of the proof of the Banach-Tarski paradox.

Among the lines of reasoning used to attempt to invalidate any proven theorem in favour of intuition, one might either try to deny the deduction, or deny the axioms on which the proof is based. For the reader that wishes to pick out those results that use the infamous Axiom of Choice (or equivalent statements), they are denoted with the conventional (**AC**).

Chapter 2

The Formation of a Paradox

There's nothing so suitable for driving a wedge into any intuition, such as between the symbolic and geometric facets of group theory, than a paradox—so below, we exhibit a proof of one.

2.1 Free groups

The production of the kind of paradox shenanigans engaged in by Hausdorff (1914), soon followed by Banach and Tarski (1924), is underpinned by the algebraic structure of free groups, which is then replicated in a far less abstract setting. The journey from group theory to geometry is not particularly difficult, but each step along the way offers more in its existence than simply being purely technical results with routine justification.

We begin with free groups, which have no immediate geometric example (compare this with, for example, the dihedral groups). An easy way to obtain a free group in the first place is to impose restrictions on the set of strings on an alphabet of letters. This works well because the set of strings over some alphabet, together with the string concatenation operation, form a monoid, and it does not take much additional work to turn it into a group.

Definition 2.1 *Let G be a finite set with n characters. Call G the set of* generators*. Let G^* denote the set of strings of finite length over the set G, and let G^{-1} denote a copy of G with they symbol $^{-1}$ appended to the right of each element, i.e. for each $x \in G$ there is a corresponding $x^{-1} \in G^{-1}$.*

Denote the empty string with the symbol e. For all $a^{-1} \in G^{-1}$, set $(a^{-1})^{-1} = a$, and for any string $\sigma = \sigma_1\sigma_2 \dots \sigma_m$, said to be of length *$m$, set $\sigma^{-1} = \sigma_m^{-1}\sigma_{m-1}^{-1} \dots \sigma_1^{-1}$.*

Let R be the operation that repeatedly removes substrings of the forms $\sigma\sigma^{-1}$ and $\sigma^{-1}\sigma$ for any $\sigma \in G$ until no such substrings remain. The set $\Sigma_G = R\left(\left(G \cup G^{-1}\right)^*\right) = \left\{R(\sigma) : \sigma \in \left(G \cup G^{-1}\right)^*\right\}$ is called the set of reduced strings over *G.*

A free group on *G, denoted F_G, is the pair $(\Sigma_G, \cdot)$, where $\cdot$ is defined by the rule $\sigma \cdot \rho = R(\sigma\rho)$, i.e. string concatenation followed by reduction. F_G is often used to refer to just Σ.*

Since $F_G \cong F_H$ if $|G| = |H| = n$, it is convenient to instead write F_n and talk about the free group of rank *n. Notation:* $\mathrm{rank}(F_G) = n$.

It would be nice if the previous definition was more succinct, and it is tempting to favour clever shortness ahead of clarity by defining it completely differently, but the above is presented the way it is for two reasons:

(i) It illustrates a methodical and precise way of obtaining a free group, which is not uncommon amongst the literature, and

(ii) Using strings, it is easier to treat the objects in the generating set abstractly, allowing great flexibility in forming free groups in unexpected places.

Most of the "heavy lifting" in Definition 2.1 that allows F_n to be a group, is contained in the action of the reduction operation R. Can the definition be made neater if there were a better way of describing the action of R?

It is worth establishing that F_G is in fact a group, and the notation $^{-1}$ is justified.

Proposition 2.2 *F_n is a group, with e as the identity and for any $\sigma \in F_n$, σ^{-1} and σ inverses of one another. For $n > 1$, it is non-commutative.*

Proof Let $\sigma, \rho, \tau \in F_n$. The operation $\cdot$ is closed, since R tidies up the result. $\sigma \cdot (\rho \cdot \tau) = \sigma \cdot R(\rho\tau) = R(\sigma\rho\tau) = R(\sigma\rho) \cdot \tau = (\sigma \cdot \rho) \cdot \tau$, so it is associative. $e \cdot \sigma = R(e\sigma) = R(\sigma) = \sigma$, so e is the identity. $\sigma \cdot \sigma^{-1} = R(\sigma\sigma^{-1}) = e = R(\sigma^{-1}\sigma) = \sigma^{-1} \cdot \sigma$, so σ and σ^{-1} are inverses.

Suppose $n = 1$. Then there is one letter, $a \in F_1$, and as shown above, a and a^{-1} commute. Since all strings in F_1 must then be composed entirely of a and a^{-1}, or is the identity, all strings in F_1 commute with one another.

Now take $n \geq 2$. Let a, b be distinct letters in F_n, and suppose $a \cdot b = b \cdot a$. Then $e = b^{-1} \cdot b = a^{-1} \cdot (b^{-1} \cdot b) \cdot a = a^{-1} \cdot b^{-1} \cdot (b \cdot a) = a^{-1} \cdot b^{-1} \cdot (a \cdot b)$. However, $a^{-1} \cdot b^{-1} \cdot (a \cdot b) = R(a^{-1}b^{-1}ab) \neq e$ since none of $a^{-1}b^{-1}$, $b^{-1}a$, or ab may be reduced. Since for $n \geq 2$ there are always at least two distinct letters, this is a contradiction to a, b commuting. □

For brevity, for positive n and string σ, let

$$\sigma^n \overset{\text{def}}{=} \underbrace{\sigma \cdot \sigma \cdots \sigma}_{n} \quad \text{and} \quad \sigma^{-n} \overset{\text{def}}{=} \underbrace{\sigma^{-1} \cdot \sigma^{-1} \cdots \sigma^{-1}}_{n},$$

and naturally, set σ^0 to be equivalent to e. For a string σ, the set $\{\sigma^n : n \in \mathbb{Z}\}$ then becomes the set of strings only involving the substrings σ or σ^{-1}.

A casual glance at F_n as described in Definition 2.1 might not glean a hint at how this helps produce a paradox as repugnant to common intuition as the infamous Banach–Tarski paradox, except perhaps that the set of reduced strings is at least countably infinite when the rank is not 0. Possibly the easiest kind of bizzare behaviour to reach for would be a "Hilbert Grand Hotel" on F_1, since $|F_1| = \aleph_0$ (Hazewinkel 2002). We can easily import most of the strange behaviour on the integers into the portfolio of strange behaviour for F_1 by the following.

Proposition 2.3 *F_1 is isomorphic to $(\mathbb{Z}, +)$, the infinite cyclic group.*

Proof Let $\left\{a, a^{-1}\right\}$ be the alphabet for F_1. Then $F_1 = \{a^n : n \in \mathbb{Z}\}$. Let $f : F_1 \to \mathbb{Z}$ be defined by the rule $f(a^n) = n$. f is then bijective.

For any a^n, a^m, $f(a^n \cdot a^m) = f(a^{n+m}) = n + m$, so f is an isomorphism. □

The similarities to cyclic groups do not hold beyond F_1. But it is precisely because F_2 and higher free groups do not behave as "nicely" as the integers that much more bizzare results follow. However, there are many natural groups that have a subgroup isomorphic to a free group. We can use group actions to leverage this.

Definition 2.4 *Let G be a group, and X be a set. A* group action *is a permutation representation of G, that is, a homomorphism Φ from G to the symmetric group on X.*

Suppose G acts on X. $\Phi(g) \in \mathrm{Sym}(X)$ is the representation of $g \in G$. The action of g on $x \in X$ *is the point $(\Phi(g))(x)$, and is denoted $g \cdot x$. For a subset $E \subseteq X$, the action of g on E is $(\Phi(g))(E) = \{(\Phi(g))(x) : x \in E\}$, and is denoted $g \cdot E$.*

An easy example is the action of matrix groups on vector spaces. Consider GL_n, the group of non-singular $n \times n$ matrices with standard matrix multiplication. The action of GL_n on an n-dimensional vector space $\mathbb{R}^n$ constitutes a homomorphism $GL_n \to \mathrm{Sym}(\mathbb{R}^n)$ as seen by the usual matrix-vector multiplication. Since it doesn't matter whether one treats matrix operations first or matrix-vector operations first—$(\mathbf{AB})\underset{\sim}{x} = \mathbf{A}(\mathbf{B}\underset{\sim}{x})$—the matrix-matrix operation is compatible with the action on the vectors. Since each invertible matrix maps vectors to vectors bijectively, they directly describe permutations on the vector space.

The notation $\cdot$ in Definition 2.4 is to remind us that there is a permutation representation of the group element, and not the group element itself, doing the work. However, a group action for a given group on a given set is often clear, so generally, spelling out the details of the permutation representation in an action doesn't contribute much to discussion. If the action of a particular element is obvious or implicit, there is no need to distinguish the element from its permutation representation, or even force awareness of the fact it is such a thing, so for $g \in G$, $x \in X$, and $E \subseteq X$, it is convenient to write the action of g on x as gx and the action of g on E as $gE = \{ge : e \in E\}$.

Note that there is a straightforward way for any group to act upon itself, namely, use the existing group operation as the action.

2.2 Paradoxicality

Definition 2.5 *Let G be a group acting on X, and $E \subseteq X$. E is said to be G-*paradoxical *or* paradoxical with respect to G *(or simply* paradoxical *if unambiguous) if, for some positive integers n, m there are pairwise disjoint subsets $A_1, \ldots, A_n, B_1, \ldots, B_m$ of E and points $g_1, \ldots, g_n, h_1, \ldots, h_m \in G$, such that $E = \bigcup_{i=1}^{n} g_i \cdot A_i$ and $E = \bigcup_{j=1}^{m} h_j \cdot B_j$.*

Definition 2.5, illustrated in Figure 2.1, says that the set E is paradoxical when it can be split up into disjoint subsets (the A_i, B_j), have each of these

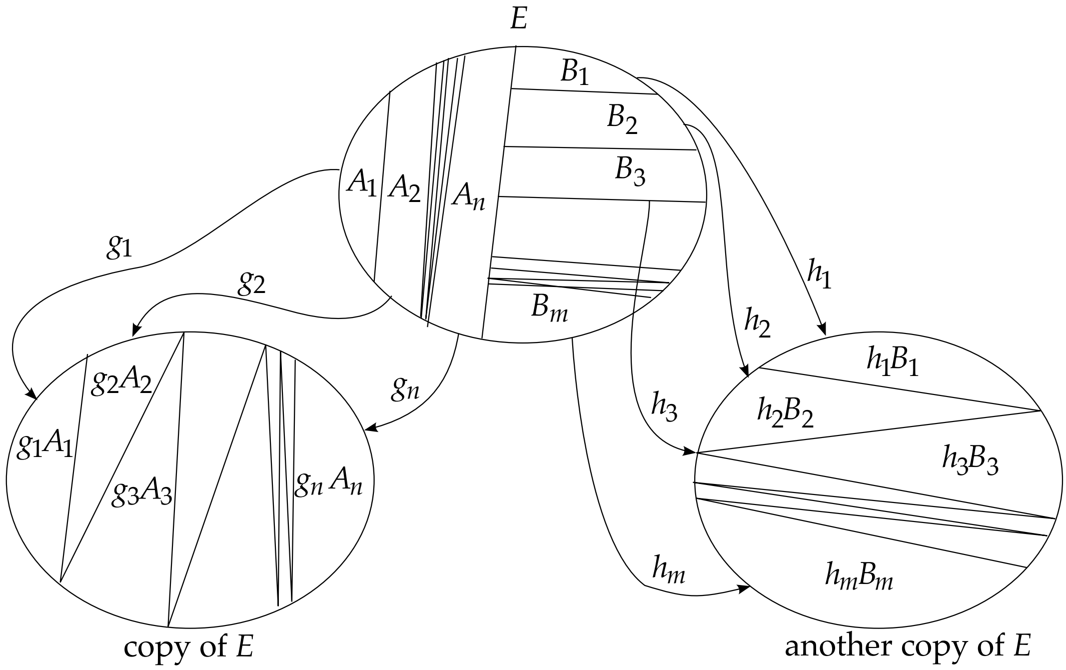

Figure 2.1: Illustration of E being paradoxical. Note that the entirety of E does not need to be used in forming the two copies.

subsets "shifted" or "rotated" by a group action, and then have some recombined into one copy of E and the others recombined into a second copy of E. Importantly, not all the subsets need to be acted upon.

Here is a simple example: the interval $(0, 1]$ can be split into disjoint subsets $(0, \frac{1}{2}]$ and $(\frac{1}{2}, 1]$. The affine transforms on $\mathbb{R}$ include scaling and translation, so $2\left(0, \frac{1}{2}\right] = (0, 1] = 2\left(\frac{1}{2}, 1\right] - 1$ uses only affine transforms. Thus $(0, 1]$ is paradoxical with respect to the affine transform group.

Definition 2.5 describes a finite number of pieces and group elements. If a countably infinite number of each is allowed, then it defines *countably paradoxical*.

For a set E to be paradoxical, at least one of the g_i and one of the h_j involved must, in some sense, "enlarge" a set. Since an action is just a permutation, an element's action to a set cannot change the cardinality of the set. The trick, then, is to make the set appear to take up more apparent space than it did before, despite there being the same number of points involved. Increasing apparent size, in general, is a soft target for creating counterintuitive behaviour in infinite sets. For example, consider translations upon on

$\mathbb{N}$. $|\mathbb{N}+3| = |\mathbb{N}|$, but $\mathbb{N}\backslash(\mathbb{N}+3) = \{0,1,2\}$, so the act of adding -3 could be said to enlarge the set $\mathbb{N}+3$, even though the cardinality is the same.

To be precise about the kind of enlargement paradoxicality aims to harness, paradoxicality aims to defeat the invariance of some measure. Suppose μ is a finitely-additive measure and is defined on each of the pieces used for E being G-paradoxical. By finite additivity of the measure on the disjoint pieces,

$$\mu(E) \geq \mu\left(\bigcup_{i=1}^{m} A_i \cup \bigcup_{j=1}^{n} B_n\right) = \sum_{i=1}^{m} \mu(A_i) + \sum_{j=1}^{n} \mu(B_n)$$

and yet also

$$\mu(E) = \mu\left(\bigcup_{i=1}^{m} g_i \cdot A_i\right) = \sum_{i=1}^{m} \mu(g_i \cdot A_i)$$

and

$$\mu(E) = \mu\left(\bigcup_{j=1}^{n} h_j \cdot B_j\right) = \sum_{j=1}^{n} \mu(h_j \cdot B_j)$$

so the action of one or more g_i and one or more h_j must increase the measure of the corresponding pieces, that is, $\mu(A_i) < \mu(g_i \cdot A_i)$ and $\mu(B_j) < \mu(h_j \cdot B_j)$.

The example of the interval $(0,1]$ with the affine transform group is a trick of infinities. The affine transforms act as bijections, which shows there are the same number of real numbers in $(0,1]$ as there are in $\left(0,\frac{1}{2}\right]$. Thus the counting measure is invariant. This is intuitively fine because we can hide behind the infinity, that is, use "$\infty = 2\infty$". An example of a measure that is not invariant is the Lebesgue measure λ, which for intervals is equal to its length, so $\lambda((0,1]) = 1$ and $\lambda\left(\left(0,\frac{1}{2}\right]\right) = \frac{1}{2}$. These values are not equal, despite being scaled versions of one another, so it is not invariant.

Intuitively, we feel there should be some groups under which λ is invariant. A space, together with a group acting on it, having a finitely-additive measure which simultaneously assigns finite values to the paradoxical subsets and yet is invariant under the group action would be quite amusing...

The following theorem provides the basis to paradoxical behaviour involving F_2 (such as the Banach–Tarski paradox), and shows that F_2 is paradoxical under its action upon itself.

Theorem 2.6 *F_2 is F_2-paradoxical.*

Proof It suffices to take a realisation of F_2: let $\Sigma = \{a, b\}$, then $F_\Sigma \cong F_2$ is a free group of rank 2. Let e denote the identity. For all (reduced) $\sigma \in F_\Sigma$, let $S(\sigma)$ be the subset of F_Σ consisting of all (reduced) strings where each starts on the left with substring σ.

As an example, a is a reduced string, and

$$\begin{aligned} S(a) &= \{a\} \cup S(a^2) \cup S(ab) \cup S(ab^{-1}) \\ &= \left\{a, a^2, ab, ab^{-1}, a^3, aba, ab^{-1}a, a^2b, \ldots\right\}. \end{aligned}$$

aa^{-1} and all strings that start with aa^{-1} aren't included because aa^{-1} is not reduced.

Now, F_Σ can be partitioned into five pairwise-disjoint subsets:

$$F_\Sigma = \{e\} \mathbin{\dot\cup} S(a) \mathbin{\dot\cup} S(b) \mathbin{\dot\cup} S(a^{-1}) \mathbin{\dot\cup} S(b^{-1}).$$

Also,

$$\begin{aligned} eS(a) \cup aS(a^{-1}) &= S(a) \cup a\left(\left\{a^{-1}\right\} \cup S(a^{-2}) \cup S(a^{-1}b) \cup S(a^{-1}b^{-1})\right) \\ &= S(a) \cup \{e\} \cup S(a^{-1}) \cup S(b) \cup S(b^{-1}) \\ &= F_\Sigma \end{aligned}$$

and by a similar argument, $eS(b) \cup bS(b^{-1}) = F_\Sigma$. Hence F_Σ is F_Σ-paradoxical with the pieces $A_1 = S(a), A_2 = S(a^{-1})$ and $B_1 = S(b), B_2 = S(b^{-1})$, and elements $g_1 = e, g_2 = a$ and $h_1 = e, h_2 = b$. □

In Theorem 2.6, only $S(a^{-1})$ and $S(b^{-1})$ required "enlargement by group action". If S is thought of as filtering out elements of F_2 by applying some restriction, then left-multiplication of $S(\sigma)$ by the inverse of the first character of σ has the effect of undoing some of the restriction (not all of it). This is depicted in the Cayley graph in Figure 2.2, which illustrates an example of the effects possible by taking cosets with S. Given the way this diagram has been presented, it also breathes life into one interpretation of "enlargement" of a set.

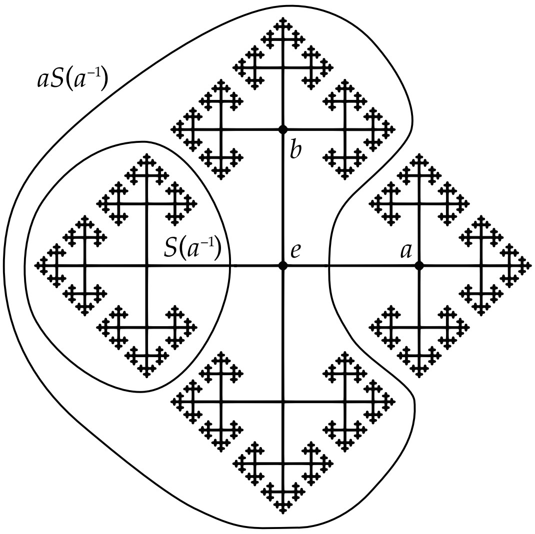

Figure 2.2: The Cayley graph for F_2, with the generators a, b and identity e labelled, and the sets $S(a^{-1})$ and $aS(a^{-1})$ circled.

2.3 Cayley graphs

The tree illustrated within Figure 2.2 deserves some explanation, though it does not form part of the proof of the paradox. It is an example of a graph designed to illustrate the structure of F_2. This requires not simply depicting the set of group elements that compose the group, but also the relationships between them.

Definition 2.7 *A* graph (V, E) *is a set of* vertices V *together with a collection of* edges $E \subseteq V \times V \times I$ *where* I *is any additional information (such as a colouring) needed to distinguish repeated pairs of vertices if necessary, otherwise* $E \subseteq V \times V$.

On paper, a graph is usually drawn as a collection of dots (one per vertex) connected by arrows or lines (one per edge). Typically, arrows are used for *directed edges*—that is, when $(u, v) \in E$ but not $(v, u) \in E$—whereas undirected lines are used for *undirected edges*—where both an edge and the opposing edge are present in E, and there is no additional distinguishing information. If there is additional information associated with vertices or edges (for example, a *labelling* $l : V \to \mathbb{N}$ or a *weighting* $w : E \to \mathbb{R}$), it is written alongside the vertex or edge it applies to. Since edge information may distinguish opposing edges (that is, it is possible that $w((u, v)) \neq w((v, u))$) such edges aren't said to be undirected, rather, there are two directed edges requiring attention.

The objective of a Cayley graph is to provide a diagrammatic insight into the structure of a discrete group by "tracing out" paths between elements in terms of generators. This is achieved as follows.

Definition 2.8 *The* Cayley graph *for a discrete group G, with generating set S, is the the graph $\Gamma(G,S) = (V,E)$ together with an* edge colouring $c : E \to S$ *given as follows: the set of vertices V bijectively corresponds (by some map $v : G \to V$) to the set of elements of G, and for all $a \in G$ and generating elements $g \in S$, there is an edge $e = (v(a), v(ag)) \in E$ having colouring $c(e) = g$.*

There are a couple of technical conventions regarding Cayley graphs. First, it is not necessary to continually distinguish a group element from the vertex representing it in a Cayley graph, so a also refers to $v(a)$. Another is that the identity is not in S—it adds nothing to the discussion. A third convention is to include all the inverses of generating elements into S. Here note that $c((a,ag)) = (c((ag,a)))^{-1}$, so every edge has an opposite but with the inverse colouring. The only undirected edges are those whose colouring is a self-inverse, since there is no information, in the form of the colouring, to distinguish direction. The other convention, not differing wildly from the previous one, is to exclude the inverses of generating elements in S, in which case, edges whose colour is not a self-inverse have no opposing edge. This convention is easier to draw, since there are less edges, but requires undirected paths—and defining $c((a,ag)) \stackrel{\text{def}}{=} (c((ag,a)))^{-1}$ in cases where the reverse edge is missing, rather than simply noting it—to uncover useful information.

Looking at the Cayley graph, it is possible to recover the sequence of generator multiplications required to move between points in the group. The vertices $v_1, v_2, \ldots, v_n$ form a *directed path* if $(v_1,v_2), (v_2,v_3), \ldots, (v_{n-1},v_n) \in E$, and they form an *undirected path* (or simply *path*) if they are a directed path or some number of edge reversals are required to make them a directed path (for example, $(v_{k+1}, v_k) \in E$ but where $(v_k, v_{k+1}) \notin E$ is needed for a directed path). Each path now corresponds to a sequence of multiplications by generating elements and their inverses. The multiplication sequence is given by the colouring function: if $v_1, \ldots, v_n$ is a path then $c((v_1,v_2)), \ldots, c((v_{n-1},v_n))$

provides the $n-1$ multiplications.

The structure of the group (modulo the given set of generators) is therefore encoded in its Cayley graph. Two groups are isomorphic when their Cayley graphs are isomorphic. However, using this fact is inefficient for computationally checking group isomorphism in general, as graph isomorphism-checking is a computationally hard problem in the class *NP* (Mulzer and Rote 2008). The converse isn't true: by picking an alternate generating set S—for instance, with redundant generators—the Cayley graph for a given group could be drastically different. Also, the presence of cycles indicates relations in the group (modulo the generating set). F_2 has no cycles in its Cayley graph, being a tree, specifically, four 3-trees glued together such that the resulting graph is still a tree.

The colouring does not necessarily have to be represented on paper using actual colours. For a free group such as F_2, generated by a couple of letters a, b, there is a nice layout of the Cayley graph on paper that obviates colours and arrowheads. Place the vertex for xa to the right of x's vertex on the page, and the vertex for xa^{-1} to the left. For b, do the same but in the vertical axis (b takes an element upwards on the page, and b^{-1} downward). Page direction now implies the direction of the edges, so arrowheads may be omitted. Figure 2.2 uses this convention.

Now Theorem 2.6 showed that F_2 has a paradoxical action on itself. For the moment consider any group, which necessarily acts upon itself by its own operation. This can be translated into an action upon its Cayley graph. Specifically, if G is a group, each $a \in G$ maps every other element $b \in G$ to ab (i.e. the left-multiplication action), and in the same manner, the action on the Cayley graph maps $v(b)$ to $v(ab)$. Furthermore, for any $x, y \in G$ there is a unique $g \in G$ such that $gx = y$ (we say the action of G on itself is *simply transitive*), g is also the unique group element mapping $v(x)$ to $v(y)$, so G also has a simply transitive action on V. This leads to the following characterisation of Cayley graphs.

Proposition 2.9 *A graph (V, E) is a Cayley graph of a discrete group G if, and only if, the action of G upon the vertices V is simply transitive and consists of graph automorphisms. (The action of g is a graph automorphism if for every $u, v \in V$, if $(u, v) \in E$ then $(g \cdot u, g \cdot v) \in E$.)*

Since the Cayley graph of any discrete free group is a tree, it is immediate that groups acting in this manner on a tree must, conversely, be free.

More will be said about F_2, the operator S, and geometry of Cayley graphs later on.

2.4 The Hausdorff Paradox

The Hausdorff paradox is that a unit sphere in $\mathbb{R}^3$, denoted S^2, has a subset that is paradoxical under the group of rotations about the origin of $\mathbb{R}^3$, denoted SO_3. This means that a subset of an ordinary sphere can be duplicated using only rotations!

A proof of the Hausdorff paradox, and in turn the Banach–Tarski paradox, uses the fact that F_2 is paradoxical under its own action. But it requires more work than that.

The next step is identifying an instance of a free group somewhere in a group acting on $\mathbb{R}^3$. In order to make the paradoxes as spectacular as they are, it is necessary to look at groups with actions that, seemingly, cannot possibly make any given set consume a different volume than it might already have. Two groups of linear transformations with this property are the group of rotations about the origin SO_3, followed by the translation group T_3, as they are isometries.

The following proof that SO_3 has a free subgroup, originates due to Mycielski and Świerczkowski (1958). It neatly avoids transcendental irrationals and uses orthogonal axes. Many other constructions, involving alternative setups with the axes and angles, are possible—the first was given by Hausdorff.

Theorem 2.10 *SO_3 contains a subgroup isomorphic to F_2.*

Proof Let A be the rotation about the x-axis by $\arccos \frac{1}{3}$, and B be the rotation about the z-axis by the same angle, i.e.

$$A^{\pm 1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & \mp\frac{2\sqrt{2}}{3} \\ 0 & \pm\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}, \quad B^{\pm 1} = \begin{bmatrix} \frac{1}{3} & \mp\frac{2\sqrt{2}}{3} & 0 \\ \pm\frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We now need to show that no reduced string on $\left\{A, B, A^{-1}, B^{-1}\right\}$, other than the empty string, is equivalent to the identity I of the rotation group, i.e. need for all strings $\sigma \in R\left(\left\{A, B, A^{-1}, B^{-1}\right\}^*\right)$, $\sigma \equiv I \Leftrightarrow \sigma = e$.

Either a reduced string ends in $B^{\pm 1}$ or it does not. Since neither B nor B^{-1} cancels with either of A or A^{-1}, we may append one of $B^{\pm 1}$ to a string if necessary to obtain a string w always ending on the right with B or B^{-1}. (This is to compensate for the fact that $A^{\pm 1}$ does not move (1,0,0)).

By induction, it will be shown that $w(1,0,0) = (a, b\sqrt{2}, c)/3^k$ for integers a, b, c, and then it will be shown that $3 \nmid b$, so $w(1,0,0) \neq (1,0,0)$, so it cannot be the identity rotation.

If w has length 1, then $w = B^{\pm 1}$ and $w(1,0,0) = (1, \pm 2\sqrt{2}, 0)/3$, so $a = 1, b = \pm 2, c = 0$.

Suppose w has length k, with either $w = A^{\pm 1}w'$ or $w = B^{\pm 1}w'$ and assume $w'(1,0,0) = (a', b'\sqrt{2}, c')/3^{k-1}$. Then $w(1,0,0) = (a, b\sqrt{2}, c)/3^k$ if $a = 3a', b = b' \mp 2c', c = c' \pm 4b'$ for $w = A^{\pm 1}w'$, or, $a = a' \mp 4b', b = b' \pm 2a', c = 3c'$ for $w = B^{\pm 1}w'$. It follows that a, b, c are integers.

It remains to be shown that $3 \nmid b$. Either w equals $A^{\pm 1}B^{\pm 1}v, A^{\pm 1}A^{\pm 1}v, B^{\pm 1}A^{\pm 1}v$, or $B^{\pm 1}B^{\pm 1}v$, for some possibly empty reduced string v. In the first and third cases, $b = b' \pm 2a'$ where $3|a'$, or, $b = b' \mp 2c'$ where $3|c'$, so if $3 \nmid b'$ then $3 \nmid b$. For the remaining two cases, let a'', b'', c'' be given by $v(1,0,0) = (a'', b''\sqrt{2}, c'')/3^l$. In either case, $b = 2b' - 9b''$, which also implies that if $3 \nmid b'$ then $3 \nmid b$. □

The core idea above is to show that no non-trivial combination of generators fixes the point $(1,0,0)$. But in reality many points will not be fixed. Consider any rotation about a line through the origin. If it is not the identity, then all but two points on the sphere—those common to the sphere and the axis of rotation—will not be fixed.

Having no non-trivial fixed points at all is a necessary condition for transferring the paradoxical action of some group on itself over to a copy of it acting on something else. The presence of some non-trivial fixed points, such as above, can be worked around.

Note that a lack of non-trivial fixed points is equivalent to having all stabiliser subgroups being trivial, since the stabiliser subgroup of G, of a given point x—written $\mathrm{Stab}_G(x)$—is, by definition, those group elements with fixed action on x.

Proposition (AC) 2.11 *Let G be a G-paradoxical group. If G acts on X and for all $x \in X$ the stabiliser subgroup is trivial, that is, $\{g \in G : gx = x\} = \{e\}$, then X is G-paradoxical.*

Proof Suppose $A_1, \dots, A_n, B_1, \dots, B_m$ are pairwise disjoint subsets of G and $g_1, \dots, g_n, h_1, \dots, h_m \in G$ satisfy $G = \bigcup_{i=1}^n g_i A_i = \bigcup_{j=1}^m h_j B_j$. By the Axiom of Choice, there is some set M containing precisely one element from each G-orbit in X, that is, each distinct Gx for $x \in X$.

Thus it is easy to see that $\{gM : g \in G\}$ partitions X as follows. Each distinct $m_1, m_2 \in M$ represents different orbits, so m_1 can't be shifted by the action of G onto m_2. Additionally, if there were some distinct $g, h \in G$ such that $gx = hx$, then $e \neq h^{-1}g \in \mathrm{Stab}_G(x)$, contradicting the hypothesis that the stabilisers are trivial. Thus for each point in $x \in X$, there is one unique $g \in G$ and one unique $m \in M$ such that $gm = x$.

Let $A_i^* = \bigcup \{gM : g \in A_i\}$ for each i, and $B_j^* = \bigcup \left\{gM : g \in B_j\right\}$ for each j. All the A_i^* and B_j^* are then pairwise disjoint. It follows that

$$\begin{aligned}
\bigcup_{i=1}^n g_i A_i^* &= \bigcup_{i=1}^n g_i \bigcup \{gM : g \in A_i\} \\
&= \bigcup_{i=1}^n g_i \bigcup_{g \in A_i} gM \\
&= \bigcup \left\{ gM : g \in \left(\bigcup_{i=1}^n g_i A_i \right) \right\} \\
&= \bigcup \{gM : g \in G\} \\
&= X,
\end{aligned}$$

and a similar argument holds for the B_j, h_j. □

To attain the Hausdorff paradox, the last step is to justify using the previous result on some part of the unit sphere that excludes fixed points. What are the fixed points? Call the set of all fixed points D. Can they simply be removed to allow the paradox to work?

Theorem (AC) 2.12 (Hausdorff Paradox) *There is a countable subset D of S^2 such that $S^2 \backslash D$ is $F_{\{A,B\}}$-paradoxical, and hence SO_3-paradoxical.*

Proof Let D be the set of all points in S^2 that are fixed under any rotation in $F_{\{A,B\}}$ (A, B rotations as in Theorem 2.10). Each $g \in F_{\{A,B\}}$ fixes two points on S^2. F_2 is countable, therefore there are at most countably many fixed points in S^2, so D is countable.

Suppose $P \in S^2 \backslash D$. If $gP \in D$ for some $g \in F_{\{A,B\}}$ then some $h \in F_{\{A,B\}}$ fixes gP, but then $g^{-1}hgP = P$, so P was fixed all along—contradiction. Thus $F_{\{A,B\}}(S^2 \backslash D) \subseteq S^2 \backslash D$, and hence Proposition 2.11 may be applied. □

2.5 The Banach–Tarski Paradox

Isn't the Hausdorff paradox taking it far enough already? For the purposes of preventing the existence of a rotation-invariant finitely-additive measure on all subsets of $\mathbb{R}^3$, it is. In this respect, the full force of the Banach–Tarski paradox could be considered extravagant and unnecessary. But such extravagance is fun. It also shows off some more tools that might go into the mischief-maker's arsenal.

The primary tool in this step is *equidecomposability*, a group action-based version of the geometric idea "congruence by dissection." Figure 2.3 illustrates equidecomposability.

Definition 2.13 *Let G act on X, and A and B be subsets of X. A and B are called* G-equidecomposable *if, for some finite n, there are subsets $A_1, \ldots, A_n \subseteq A$ and $B_1, \ldots, B_n \subseteq B$ such that*

$$A = \bigcup_{i=1}^{n} A_i, B = \bigcup_{i=1}^{n} B_i$$

where the A_i are pairwise disjoint, the B_i are pairwise disjoint, and there are group elements $g_1, g_2, \ldots, g_n \in G$ such that $g_i A_i = B_i$ for all $i = 1, \ldots, n$.

Notation: $A \sim_G B$

Synonyms: finitely G-equidecomposable, piecewise G-congruent

$\sim_G$ is straightforwardly shown to be an equivalence relation. Reflexivity and symmetry are obvious: $A \sim_G A$ since $A = 1A$, and $X \sim_G Y \Rightarrow Y \sim_G X$ since every piece in Y may be acted on by the inverse of the element required to obtain it from the corresponding piece in X (i.e. $Y_i = g_i X_i$ so $X_i = g_i^{-1} Y_i$).

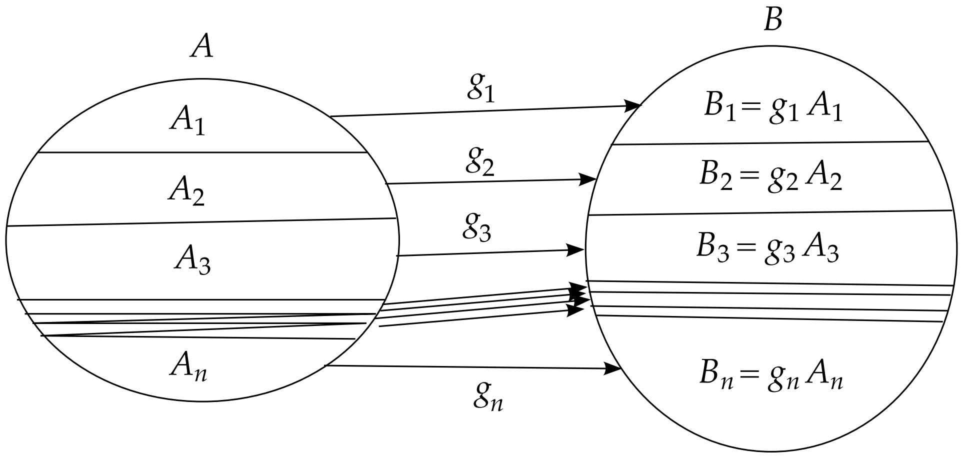

Figure 2.3: Illustration of A and B being equidecomposable with n pieces.

Showing transitivity is fairly routine. If $A \sim_G B$ in m pieces and $B \sim_G C$ in n pieces, then $A \sim_G C$ in as many as mn pieces:

Let $A_1, \ldots, A_m \subseteq A$, $B_1, \ldots, B_m \subseteq B$ and $g_1, \ldots, g_m \in G$ witness $A \sim_G B$, and likewise $B'_1, \ldots, B'_n \subseteq B$, $C_1, \ldots, C_n \subseteq C$ and $h_1, \ldots, h_n \in G$ witness $B \sim_G C$.

Let $\hat{B}_{jk} = B_j \cap B'_k$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Then take $\hat{A}_{jk} = g_j^{-1}\hat{B}_{jk}$ and $\hat{C}_{jk} = h_k\hat{B}_{jk}$. Hence the $\hat{A}_{jk}, \hat{C}_{jk}$ and $h_k g_j$ witness $A \sim_G C$.

So what is the point of $\sim_G$? It provides an alternative definition of paradoxicality: E is G-paradoxical if and only if $\exists A, B \subseteq E$ disjoint such that $A \sim_G E \sim_G B$. This is shown in Figure 2.4.

Also, it can be used to transfer paradoxicality around different sets in the same $\sim_G$-equivalence class. This implies that paradoxicality is a property of equivalence classes, and not particular to a specific set.

Proposition 2.14 *Let G act on X, let $E, E' \subseteq X$ and suppose $E \sim_G E'$. If E is G-paradoxical, so is E'.*

Proof Note the property of $\sim_G$ that if $E \sim_G E'$ then every subset $F \subseteq E$ has a corresponding $F' \subseteq E'$ such that $F \sim_G F'$ (via intersection with the pieces used for $E \sim_G E'$). Thus for the disjoint $A, B \subseteq E$ witnessing E being G-paradoxical, there are $A', B' \subseteq E'$ such that $A \sim_G A'$ and $B \sim_G B'$. Therefore $E' \sim_G E \sim_G A \sim_G A'$ and $E' \sim_G E \sim_G B \sim_G B'$.

It remains to be shown that $A' \cap B' = \emptyset$. Consider the pieces and group elements $E_1, E_2, \ldots, E_n$ and $g_1, g_2, \ldots, g_n$ such that $E'_1 = g_1E_1, E'_2 = g_2E_2, \ldots, E'_n =$

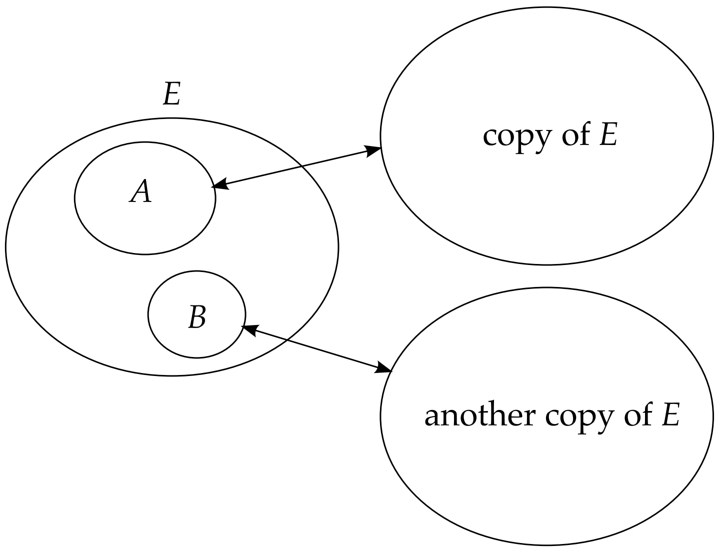

Figure 2.4: Illustration of paradoxicality of E via equidecomposibility.

$g_n E_n$, i.e. the paraphernalia witnessing $E \sim_G E'$. Take any piece E_i. A and B are disjoint, so $A \cap E_i$ and $B \cap E_i$ are disjoint, thus since each g_i acts as a bijection on X, $g_i(A \cap E_i) = A' \cap E_i'$ is disjoint from $g_i(B \cap E_i) = B' \cap E_i'$, and hence A' and B' are disjoint. □

If E is G-paradoxical, there are disjoint subsets A, B equidecomposable to E, so, by 2.14, A and B are in turn paradoxical, so they have disjoint subsets $A_A, B_A \subseteq A$ and $A_B, B_B \subseteq B$ such that $A \sim_G A_A, B_A$ and $B \sim_G A_B, B_B$, and again for these new subsets, *ad infinitum*. This is a recursive self-similarity, which, as illustrated in Figure 2.5, resembles the famous Cantor set. Note however that each layer need not remove points from the set, as $A, E \backslash A$ are two disjoint sets that might validly show E to be paradoxical. Hence it is necessary for all groups that cause a set to be paradoxical to induce this self-similarity.

As for the matter of the paradox, all that remains is to show $S^2 \sim_{SO_3} S^2 \backslash D$, and then—finally—$S^2$ is paradoxical by virtue of being in the same class as $S^2 \backslash D$. But how?

There needs to be an easier test for equidecomposibility. Not only that, it would be nice to be able to *absorb* misbehaving points.

In yet another stroke of co-opting results from other branches of mathematics, Banach showed that a generalisation of the Cantor–Schröder–Bernstein

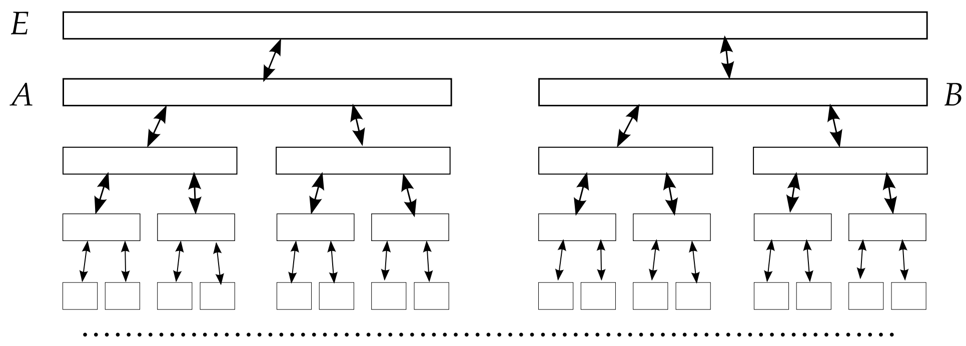

Figure 2.5: A paradoxical set E, its paradoxical subsets A and B, and so on. All are equidecomposable to one another, and have the same cardinality.

theorem[1] would not only have the mechanics to do even better than this, but could be extended to a partial order based on any equivalence relation ρ satisfying two conditions, namely,

1. if $A \,\rho\, B$ then there is a bijection $b : A \to B$ such that $C \,\rho\, b(C)$ whenever $C \subseteq A$, and
2. if $A_1 \cap A_2 = \emptyset = B_1 \cap B_2$, and if $A_1 \,\rho\, B_1$ and $A_2 \,\rho\, B_2$, then $(A_1 \cup A_2) \,\rho\, (B_1 \cup B_2)$.

The cardinal equality relation in the classical Cantor–Schröder–Bernstein theorem, i.e. $|A| = |B|$, satisfies these conditions, and so too does $\sim_G$. In the classical case of course $|A| \leq |B|$ is defined by the existence of some injection from A to B, but the proof need not be done in those terms.

If $\preceq$ is defined by the relation $A \preceq B \Leftrightarrow A \,\rho\, C$ for some $C \subseteq B$, then it is reflexive and transitive: $A \subseteq A$ and $A \,\rho\, A$ so $A \preceq A$, and if $A \preceq B$ and $B \preceq C$ then $\exists D \subseteq B$ such that $A \,\rho\, D$ and $\exists E \subseteq C$ such that $B \,\rho\, E$, so by the condition 1, the existence of a bijection $B \to C$, $D \,\rho\, b(D) \subseteq E \subseteq C$, and hence $A \,\rho\, b(D) \subseteq C$, giving $A \preceq C$. The remainder (and bulk) of the work—as is the case for the classical result—is in showing antisymmetry.

Theorem 2.15 (Banach–Cantor–Schröder–Bernstein Theorem) *Suppose $A, B \subseteq X$. Let ρ be an equivalence relation satisfying conditions 1 and 2 above, and let $\preceq$ be defined by the condition $A \preceq B \Leftrightarrow \exists C \subseteq B$ such that $A \,\rho\, C$.*

[1]Cantor's name has frequently been omitted, but he was in fact the first to prove the theorem, as a corollary of the well-ordering theorem (Hinkis n.d.). The proof here however does not involve the Axiom of Choice, and is the one the proof to Theorem 2.15 is based on.

If $A \preceq B$ and $B \preceq A$ then $A\,\rho\,B$ by the above, hence $\preceq$ is a partial ordering of the ρ-classes in $\mathscr{P}(X)$.

Proof Suppose $A \preceq B$ and $B \preceq A$. Let $f : A \to B_1$ and $g : A_1 \to B$ be bijections, where $B_1 \subseteq B$ and $A_1 \subseteq A$, as seen in condition 1. Let $C_0 = A \backslash A_1$ and for each $n = 0, 1, 2, \ldots$ define $C_{n+1} = g^{-1}(f(C_n))$. Finally, let $C = \bigcup_{n=0}^{\infty} C_n$.

$A\backslash C$ is the remainder of A_1 untouched by the formation of C, and $B\backslash f(C)$ is likewise the remainder of B_1 on the B side, so since g is a bijection, $g(A\backslash C) = B\backslash f(C)$. By condition 1, since $A\backslash C \subseteq A$, $A\backslash C \sim g(A\backslash C) = B\backslash f(C)$ and $f(C) \sim C$, so using condition 2, $A = (A\backslash C) \cup C \sim (B\backslash f(C)) \cup f(C) = B$. □

So how does this help finding more sets equidecomposable to a given one? Let $\preceq_G$ denote $\preceq$ as above for the equivalence relation $\sim_G$. Suppose E is some set with two disjoint subsets A and B. If $E \sim_G B$, then

- $E \sim_G B$, which is a subset of $E\backslash A$, so $E \preceq E\backslash A$, and
- $(E\backslash A) \sim_G (E\backslash A)$, a subset of E, so $E\backslash A \preceq E$.

Hence $E\backslash A \sim_G E$, and since A is only required to be disjoint from B, it can be made as large as $E\backslash B$ or as small as $\emptyset$, the essential feature being $B \subseteq E\backslash A$.

Furthermore, if $A \sim_G E$, then by the same argument, $E\backslash B \sim_G E$, so then $E\backslash B \sim_G A \sim_G E \sim_G B \sim_G E\backslash A$, i.e. everything considered is G-equidecomposable with everything else.

So finally, we approach the justification for $S^2 \sim S^2\backslash D$. The only remaining thing to note is that it is not necessary to stay confined to the free subgroup of SO_3 in order to prove SO_3-equidecomposibility.

Theorem 2.16 *If D is a countable subset of S^2, then $S^2 \sim_{SO_3} S^2\backslash D$.*

Proof Let l be a line through the origin that does not intersect D (because D is countable, there are uncountably many points in $S^2\backslash D$ that can be chosen to form such a line). Consider rotations about l: Let $\rho(\theta)$ be the rotation by θ around l. Construct a set J of angles θ that, for some integer n, take some point $P \in D$ to some other point $\rho(n\theta)P \in D$. There are countably many points in D, and so there are countably many angles that could be used to move between them. Thus J is countable.

Choose an angle from the uncountably many remaining, i.e. fix some $\theta \notin J$. By definition, $\rho(n\theta)P \notin D$ for all $n \neq 0$ and $P \in D$. Now $\rho(n\theta)D \cap D = \emptyset$, and it is clear that this entails for any distinct integers m, n that $\rho(m\theta)D \cap \rho(n\theta)D = \emptyset$.

Let $\bar{D} = D \cup \rho(\theta)D \cup \rho(2\theta)D \cup \cdots$. Now

$$\begin{aligned}\bar{D}\backslash\rho(\theta)\bar{D} &= (D \cup \rho(\theta)D \cup \rho(2\theta)D \cup \rho(3\theta)D \cup \cdots) \\ &\qquad \backslash\,(\rho(\theta)D \cup \rho(2\theta)D \cup \rho(3\theta)D \cup \cdots) \\ &= D\end{aligned}$$

so $(S^2\backslash\bar{D}) \cup \rho(\theta)\bar{D} = S^2\backslash D$.

$\bar{D} \sim_{SO_3} \rho(\theta)\bar{D}$. Hence, by condition 2 from the paragraphs before Theorem 2.15,

$$S^2 = \bar{D} \cup (S^2\backslash\bar{D}) \sim_{SO_3} \rho(\theta)\bar{D} \cup (S^2\backslash\bar{D}) = S^2\backslash D.$$

□

SO_3 is just the group of rotations. Since nothing so far uses the radius of a sphere, it can all be generalised to spheres of any radius—or all radiuses, either up to some finite value for a ball, or beyond, simultaneously. Furthermore, if translations are also allowed—they are isometries too, so a paradox on this larger group G_3 is just as incredible—then spheres and balls in other locations also work.

Corollary (AC) 2.17 (The Banach–Tarski Paradox) *S^2 is SO_3-paradoxical, and any sphere, and any solid ball centred at the origin is SO_3-paradoxical. Furthermore, any sphere or solid ball in $\mathbb{R}^3$—and $\mathbb{R}^3$ itself—are G_3-paradoxical.*

Proof The Hausdorff paradox in Theorem (AC) 2.12 gave $S^2\backslash D$ to be SO_3-paradoxical for some countable subset D. Theorem 2.16 gives $S^2\backslash D \sim_{SO_3} S^2$, so by Proposition 2.14, it can be transferred over to S^2.

Inside G_3, which contains the group of translations, spheres and balls located anywhere may be relocated to be centred at the origin, so without loss of generality, assume that the sphere/ball is a unit sphere or ball at the origin. To extend the paradoxicality of the sphere to the unit ball without its centre $B\backslash\{\mathbf{0}\}$, associate with every $P \in S^2$ the ray $\{\alpha P : 0 < \alpha \leq 1\}$. Now the paradoxicality of S^2 is straightforwardly applied to the rays. For $\mathbb{R}^3 \backslash \mathbf{0}$, perform the same trick but without bounding above the α.

The remaining issue is the origin. Let $P = (0, 0, \frac{1}{2})$, let l be a line through P but not through $\mathbf{0}$, and let ρ be a rotation of infinite order (such as, say, $\arccos \frac{1}{3}$ radians) about l. Repeating essentially the same trick as in Theorem 2.16: let $D = \{\rho^n(\mathbf{0}) : n \geq 0\}$, then $\rho(D) = D \backslash \{\mathbf{0}\}$, so $D \sim_{G_3} D \backslash \{\mathbf{0}\}$, and thus $B = D \cup (B \backslash D) \sim_{G_3} \rho(D) \cup (B \backslash D) = B \backslash \{\mathbf{0}\}$. □

But why stop *there*?

Corollary (AC) 2.18 (The Strong Banach–Tarski Paradox) *If A, B are any two bounded subsets of $\mathbb{R}^3$, each with non-empty interior, then $A \sim_{G_3} B$.*

Proof Let K, L be solid balls such that $A \subseteq K$, but $L \subseteq B$, and let n be sufficiently large to allow K to be covered (overlaps permittied) by n copies of L. By the Banach–Tarski paradox, L may be duplicated into n copies of L and shifted by translations so they are all disjoint—call the set of copies S. Thus $S \sim_{G_3} L$. $K \preceq S$ by translations. Therefore $A \subseteq K \preceq S \preceq L \subseteq B$, hence $A \preceq B$.

By a similar argument, $B \preceq A$, so by the Banach–Cantor–Schröder–Bernstein theorem, $A \sim_{G_3} B$. □

2.6 Responses

The Banach–Tarski paradox, in its full glory, is an astounding result. So astounding, in fact, that it takes a little while to believe some sort of fraud is not taking place. There are several objections people have raised to invalidate it or attempt to reconcile it with geometric intuition.

The first objection is to blame the Axiom of Choice. How can we trust an axiom that gives the existence of objects non-constructively? The pieces used in the Banach-Tarski paradox are surely the epitome of non-constructibility. There are problems with this objection. The first objection is that there are other strange geometric results that do not depend on the Axiom of Choice, such as the Sierpiński-Mazurkiewicz paradox.

Secondly, the strength of the Banach–Tarski paradox in relation to the Axiom of Choice has shifted over the years. While the Axiom of Choice certainly implies the Banach–Tarski paradox, the Banach–Tarski paradox has not been shown to imply the Axiom of Choice. The quest is then on to find the weakest system of which it is a theorem, and what weakened flavours of the Axiom of Choice, such as dependent choice, the Banach–Tarski paradox might entail. It was stated by Wagon (1985) that in likelihood the paradox was fully equivalent to the Axiom of Choice.

In partial answer to this, a marvellously succinct article by Pawlikowski (1991) following work by Foreman and Wehrung (1991), showed that the Banach–Tarski paradox can be proved with only the addition of the Hahn–Banach theorem—equivalent to the ultrafilter lemma—and therefore does *not* require the full Axiom of Choice. Of course, now one may try to exercise the same argument as above, "Hahn-Banach theorem" for "Axiom of Choice" *mutatis mutandis*, however such objections are even less justifiable than in the case of the Axiom of Choice since the Hahn-Banach theorem may be used to show an even larger class of useful theorems.

Finally, the Axiom of Choice is really only involved in Proposition 2.11; if there were some other way to transfer the paradoxical action of F_2 on itself to the points it acts upon, then AC can be dispensed with, but the involvement of AC at that step would hardly seem objectionable. Pawlikowski (1991) does not use the Hahn-Banach theorem to address this step of the proof specifically, rather, the paradox is implied via the use of the semigroup of equidecomposability types discussed by Wagon (1985).

The second approach to dealing with the intuition-logic incongruence, following more along the technique of *monster-barring*[2] (Lakatos 1976) is to accept it and modify the notion of measure, so that it is fine if the measure of a particular subset is allowed to change under the seemingly measure-preserving isometries.

The Hausdorff and Banach–Tarski paradoxes show (in the latter, quite extravagantly) that there cannot exist a finitely-additive measure that is isometry-invariant and with a finite total measure, on *all subsets of the space* if it is to be preserved even by rotations. In terms of decomposition, S^2 can be

[2]How ironic that monster-barring leads to monsters.

assigned a measure, and two disjoint copies of S^2 can be assigned twice that measure by finite additivity, but what of the pieces used in the equidecomposition passing between them? Must they have a defined measure? Let's exclude them. As a geometric justification for this, note that since the orbits of the F_2 rotation group used are (intuitively) disconnected points scattered across the sphere, and the resulting set of representatives M is created with the Axiom of Choice, the pieces are hardly "pieces" in the natural sense of the word. The sets would require at least infinite precision to "cut". So to accept that the required pieces have undefined measure, and that it would be possible to add them together in multiple ways to get sets with defined but differing measures, now looks quite natural. Note that once again, the full Axiom of Choice is not required to produce non-Lebesgue measurable sets (Foreman and Wehrung 1991).

Such non-measurable sets (*Vitali sets*, in honour of the first to formulate some examples) are then simply excluded from the collection of measurable sets. Can the non-measurable sets used in the paradoxes be characterised? Attempting to answer this turns out to be a most fruitful approach.

The misbehaving object becomes the copy of F_2 inside SO_3, since without the magical property F_2 seems to have, such bizzare pieces of a sphere would be impossible to consider. von Neumann (1929) made the observation, which, to be more precise, is: whether or not finitely-additive isometry-invariant measures with finite total measure exist is a property of the group in question. This led to a very neat direction for group-theoretic research following the publishing of the Banach–Tarski paradox. Let NF be the class of all groups that do not contain a free subgroup (of order 2) and let AG be the class of groups where there exists a finitely-additive group-action-invariant measure of finite total measure (the *amenable* groups) on the group. What relation do NF and AG have? All amenable groups cannot not have a free subgroup (by the paradoxes), so $AG \subseteq NF$.

Are groups without free subgroups the only amenable groups? The problem, known as the von Neumann conjecture, was answered in the negative only as recently as 1979 by A. Yu. Ol'shanskii—after decades of important work, such as the development of van Kempen's lemma, small cancellation theory, and other concepts in geometric group theory. Ol'shanskii's counterexample to the von Neumann conjecture is known as a *Tarski monster*.

Chapter 3

Groups, Generators, Graphs and Geometry

A brief exploration of fundamental groups of pointed topological spaces[1] and—turning the chessboard over—van Kempen's Lemma. Much of the contents of this chapter forms the basis for producing the non-amenable periodic Tarski monster of Ol'shanskii (1991).

3.1 Surfaces

Definition 3.1 *A topological space X is called* path-connected *if every pair of points $a, b \in X$ there is a continuous* path $p : [0,1] \to X$ *having* $p(0) = a, p(1) = b$.

Path-connectedness is a stronger condition than "vanilla" connectedness, which is simply the non-existence of two disjoint open sets whose union is the entire space.

Proposition 3.2 *Every path-connected space $(X, \mathscr{T})$ is connected.*

Proof Suppose $(X, \mathscr{T})$ is path-connected but that $X = A \cup B$ for disjoint $A, B \in \mathscr{T}$. Let p be a path connecting points $a \in A, b \in B$. p is continuous, so $\overleftarrow{p}(A)$ and $\overleftarrow{p}(B)$ (inverse images) are also open in $[0,1]$. Now $[0,1] = \overleftarrow{p}(A) \cup \overleftarrow{p}(B)$ as A, B together include the image of p, and $\overleftarrow{p}(A)$, $\overleftarrow{p}(B)$ must be

[1] Point-set topological spaces as opposed to, of course, pointless ones.

disjoint. But since every interval on the real line—and in particular, $[0, 1]$—is connected, this is a contradiction. Hence path-connectedness implies connectedness. □

Now, to define *surface* in a useful manner for discussing diagrams, it is necessary to introduce triangulations and cell decompositions.

A *topological triangle* in a space $(X, \mathscr{T})$ is just an *embedding* (continuous injection) of a traditional triangle from $\mathbb{R}^2$ into that space. A topological triangle doesn't necessarily have all the properties of a traditional triangle[2] however an important property is that the edges of a topological triangle don't intersect one another. Another property is that the interior of the triangle is homeomorphic to a disc, so the boundary of a triangle is essentially an embedding of a circle with three distinguished points (the vertices). Similarly, a topological n-gon is an embedding of a disc with n distinguished vertices on its boundary.

A *triangulation* Δ of a topological space $(X, \mathscr{T})$ is a finite set of topological triangles where $X = \bigcup \Delta$ and the intersection of any two triangles is either completely empty, a single shared vertex, or a single shared edge. A space that admits a triangulation might be thought of as taking n triangles and gluing them together, so since there are finitely many triangles, each of which is compact, the space is compact. Furthermore, any space admitting a triangulation is Hausdorff (the T_2 separation axiom).

Consider a vertex v from a triangulation Δ of a space. The *star* at v, denoted $\mathrm{St}(v)$, is given by $\mathrm{St}(v) = \{T \in \Delta : v \in T\}$—the set of all triangles sharing that vertex. There are various kinds of star, and for our purposes we will need two: circular and semi-circular, defined as follows.

Let $e_1, e_2, \ldots, e_n$ be the edges of triangles in $\mathrm{St}(v)$ that have v as a vertex. If there are n triangles $T_1, T_2, \ldots, T_{n-1}$ in $\mathrm{St}(v)$ such that e_1, e_2 are edges of T_1, e_2, e_3 edges of T_2, $\ldots$, e_{n-1}, e_n edges of T_{n-1}, and e_n, e_1 edges of T_n, the star is *circular*.

If, instead, there are $n - 1$ triangles $T_1, \ldots T_{n-1}$ in $\mathrm{St}(v)$ such that e_1, e_2 are edges of T_1, $\ldots$, and e_{n-1}, e_n are edges of T_{n-1}, then the star is *semi-circular*. These two kinds of star are depicted in Figure 3.1.

[2]For example, internal angles summing to π—the notion of angle might be completely nonsensical in the space!

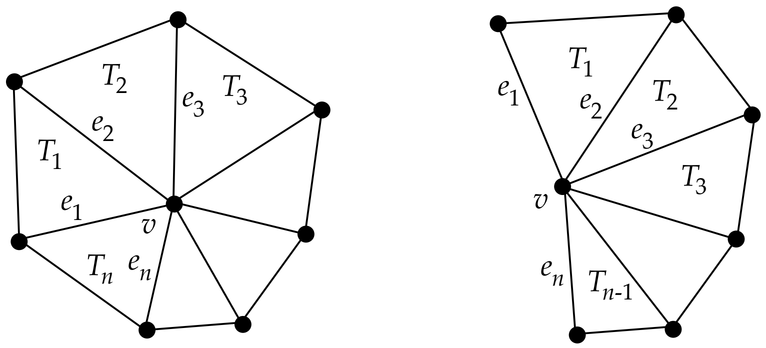

Figure 3.1: Stars characterising a surface. Left: a circular star at v. Right: a semicircular star, with edges e_1 and e_n being boundary edges.

Now there is sufficient mechanics for stating the following definition.

Definition 3.3 *A connected space* $(X, \mathscr{T})$ *is a* surface *if it admits a triangulation where every star is either circular or semicircular.*

Examples of surfaces are plentiful: path-connected regions in the plane $\mathbb{R}^2$, the unit sphere S^2, the ordinary torus T^2, the surface of an ordinary icosahedron, and the Möbius strip are all surfaces. Surfaces where all stars are circular are *closed*, otherwise they have a *boundary* ∂X, given by the edges each belonging to only one triangle from the semi-circular stars. Note that an edge cannot be shared by more than two triangles in a triangulation of a surface.

3.2 Cell decompositions

Cell decompositions are similar to triangulations. The basic difference is that the unit of slicing up the surface is a topological n-gon (where n is not fixed) rather than a 3-gon (triangle). The embedded n-gons in a surface are called *n-cells*. The n-cell differs from the topological triangle in that n is arbitrary and the vertices and sides are identified by a continuous mapping from an n-gon in the plane.

Definition 3.4 *Let* P *be an* n*-gon in the plane, and* $f : P \to X$ *a continuous mapping into the space* $(X, \mathscr{T})$*. If*

(i) *the restriction of f to the interior of P is an embedding,*

(ii) *the interior of each edge of P are embeddings in X*

(iii) *for $a, b \in P$, $f(a) = f(b)$ implies a and b are in the boundary of P, and if a is a vertex so is b, otherwise the edge containing a coincides with the edge containing b,*

then (Π, f) *for* $\Pi = f(P)$ *is an* n-angular cell *or simply* n-cell.

Thus a *cell decomposition* is a finite set of cells $\{(\Pi_i, f_i)\}_{i=1}^m$ with a similar conditions to a triangulation, namely, $X = \bigcup_{i=1}^m \Pi_i$, and the intersection of pairs of cells is either empty or limited to shared vertices and shared edges. Stated another way, a triangulation is a cell decomposition where all embedded n-gons are triangles.

There is a sense in which cell decompositions are coarser than triangulations. A cell decomposition has some number of *refinements*, where a refinement has more cells, each of which is contained in a cell from the original decomposition, and all the edges and vertices from the original decomposition appear in the refined decomposition. This might be thought of as adding additional edges to the decomposition that do not cross any existing ones. By applying a subdivision in this manner to each of the cells, a refinement which is a triangulation can be obtained.

3.2.1 Orientations

Each edge of a cell may be directed, similarly to a directed graph. If all the edge directions of a cell are head-to-tail, as in Figure 3.2, then the cell has an *orientation*. A cell orientation can be thought of as choosing either clockwise or anticlockwise “flow” for a cell. This is depicted in diagrams with arrowheads indicating direction around the cell. In a cell decomposition, the edges not on a boundary are shared by two cells, so they may be given one or other of the possible directions, or both. When there is a choice of orientation for each cell such that all the non-boundary edges are assigned both opposing orientations, the surface is said to be *orientable*.

Orientable surfaces include the closed unit square, the sphere, the annulus, and the torus. Non-orientable surfaces include the Möbius strip and the Klein bottle.

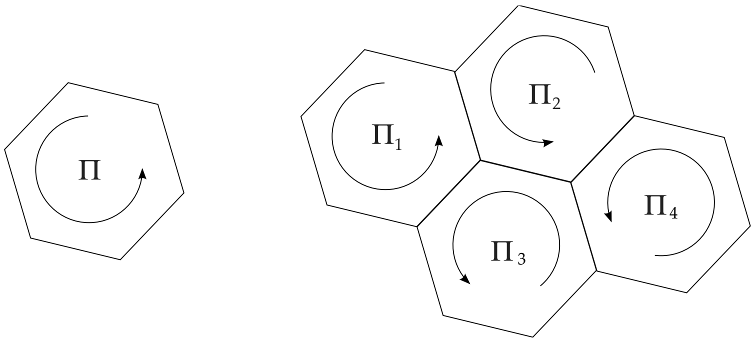

Figure 3.2: Left: a 6-cell with orientation. Right: part of an orientable cell decomposition.

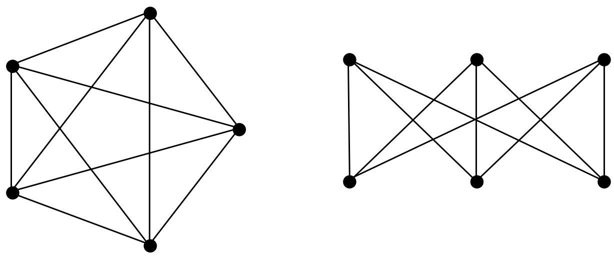

Figure 3.3: The graphs whose subdivisions characterise non-planar graphs in Kuratowski's Theorem, K_5 (left) and $K_{3,3}$.

3.3 Topological graphs

We reconsider briefly the act of drawing a graph (Definition 2.7) on paper, and consider extending this to a more general surface.

It would be nice if drawing a graph did not require intersecting edges, that is, for all the edges to be drawn as continuous paths by some injective function. If such a drawing can be made for a graph on the plane, then it is said to be *planar*. A theorem of Kuratowski (1930) holds that any finite graph is planar if and only if it excludes subdivisions of the graphs K_5 and $K_{3,3}$, illustrated in Figure 3.3, as subgraphs. K_5 is the complete graph of 5 vertices, *complete* meaning each vertex is connected directly to each other vertex. $K_{3,3}$ is the complete *bipartite* graph of two sets of three vertices—all three vertices in one subset are connected to each of the other three vertices.

There are a variety of interesting theorems regarding graph planarity, and a few linear-in-$|V|$-time algorithms for determining planarity of finite graphs. A modern example of such an efficient algorithm is given by Boyer and Myrvold (2004).

The notion of planarity extends to other surfaces. A graph may be *embedded* into some surface if it can be drawn on the surface without intersecting edges. If a graph is planar, then it can be embedded into the plane. Graphs that are not planar might be embeddable in surfaces other than the plane, for example, K_5 can be embedded into a torus, but not the plane. K_5 is thus a *toroidal* graph.

Note that each cell decomposition of a surface provides an embedded topological graph, and every embedded graph with edges covering the entire boundary of the surface (if there is one) provide a cell decomposition.

We may also treat a graph as a topological space on its own, without embedding it in some other space. This is done as follows. Each edge is treated as being homeomorphic to the unit interval $[0, 1]$. The vertices at either end of an edge provide the appropriate gluing points, but no other points are considered equivalent. Finally, regardless of the planarity or embeddability of the graph in a surface, there are no faces—the interior of what might be faces are not in the topology at all. This endows any arbitrary graph with a topology regardless of what surfaces it might be embeddable in.

Some basic questions about the graph topology described above arise: is it metrizable? Yes. One can first define a metric d along each edge to be a (largely arbitrary) distance between the two end vertices taking values between 0 and 1, since each edge is homeomorphic to $[0, 1]$. Then, for vertices, $d(u, v)$ will be the number of edges in the path connecting u to v. For pairs with multiple paths connecting them, the metric is well-defined if the shortest path is always chosen.

Now consider the single-vertex, single-edge directed graph in Figure 3.4. It is a *loop* or *circle*. In terms of the homeomorphism from $[0, 1]$, $f_a(0) = f_a(1)$. It is homeomorphic to a circle in the plane, and has one distinguished point, making it a 1-cell when embedded in a surface.

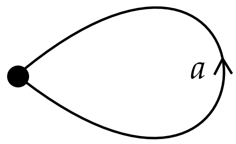

Figure 3.4: The graph with one edge a and one vertex.

3.4 Fundamental groups

We can define a structure based on travelling around loops, such as the one in Figure 3.4, starting from some fixed base point in a path-connected topological space X as follows.

Consider the family L_X of all possible continuous loops in the space starting and ending with an arbitrary point x_0—the *base point*. For a loop $a \in L_X$, consider the motion of travelling once around a in the positive direction (illustrated diagrammatically by an arrowhead). The path for a is traced out by a continuous $f_a : [0,1] \to X$, and since a starts and ends with x_0, $x_0 = f_a(0) = f_a(1)$.

Let e denote an "identity" loop consisting of only the base point, or, in terms of motion, e means "no travel." As constant functions are continuous, $e \in L_X$.

Let a^{-1} denote the loop having the reverse path, that is, $f_{a^{-1}}(x) = f_a(1-x)$. Alternatively, a^{-1} has the path taken by travelling around a in the reverse direction. Thus $a^{-1} \in L_X$.

For loops $a, b \in L_X$, define $a \cdot b$ to be the loop given by the path

$$f_{a\cdot b}(x) \stackrel{\text{def}}{=} \begin{cases} f_a(2x) & 0 \le x \le \frac{1}{2} \\ f_b(2x-1) & \frac{1}{2} \le x \le 1 \end{cases} .$$

Thus $a \cdot b$ is the combined loop formed by travelling along a and then along b.

Now the paths for $a^{-1} \cdot a$ and $a \cdot a^{-1}$ can't yet be thought of as being equal to e, but this will be fixed as follows.

Define an equivalence relation ψ on L_X by setting $a \,\psi\, b$ if and only if there exists a continuous $H : [0,1] \times [0,1] \to X$ such that $H(0,x) = f_a(x)$ and $H(1,x) = f_b(x)$ holds for all $x \in [0,1]$. This can seem like an eyeful, but the core idea is *continuous deformation*. Where it exists, the function H is called a *homotopy*, and describes a way to continuously deform a into b without shifting the base point. Thus we are interested in the ψ-equivalence classes (*homotopy classes*), which consist of loops continuously deformable to one another.

Definition 3.5 *The* fundamental group *of a path-connected topological space X, denoted $\pi_1(X)$, is the set of homotopy classes L_X/ψ together with the operation $\cdot$ described above (extended to the homotopy classes).*

To avoid the use of additional notation, loops will be considered defined only up to homotopy equivalence, that is, $a = b$ will mean $a \,\psi\, b$.

The use of homotopy is important for a variety of reasons. Consider associativity. In $(a \cdot b) \cdot c$,

$$f_{(a\cdot b)\cdot c}(x) = \begin{cases} f_a(4x) & 0 \le x \le \frac{1}{4} \\ f_b(4x-1) & \frac{1}{4} \le x \le \frac{1}{2} \\ f_c(2x-1) & \frac{1}{2} \le x \le 1 \end{cases}$$

which is equivalent under a continuous deformation to

$$f_{a\cdot(b\cdot c)}(x) = \begin{cases} f_a(2x) & 0 \le x \le \frac{1}{2} \\ f_b(4x-2) & \frac{1}{2} \le x \le \frac{3}{4} \\ f_c(4x-3) & \frac{3}{4} \le x \le 1 \end{cases}$$

being the description for $a \cdot (b \cdot c)$.

Homotopy classes also fixes the problem with inverses: $a^{-1} \cdot a$ is homotopic to a point, since the interior of the loop $a^{-1} \cdot a$ necessarily does not have any punctures or other problematic points on its (empty) interior, so there is nothing that would force all deformations of $a^{-1} \cdot a$ to a point to be discontinuous.

Let's consider the usual suspects. The fundamental group of the plane is trivial group since all loops can be continuously deformed into a point. Loops on a sphere are also continuously deformable to a point, so it too has a trivial fundamental group. Up to homotopy equivalence, the plane with a puncture has one loop (which can't be continuously deformed across

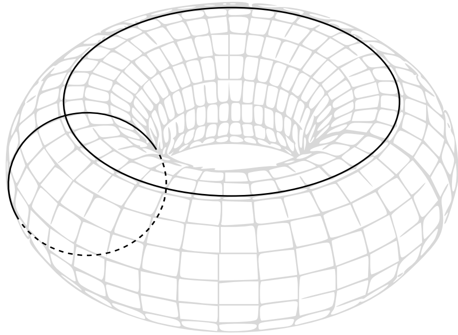

Figure 3.5: A 1-torus with one loop highlighted from each homotopy class.

the puncture—a deformation must "leap across") and so, it has the same fundamental group as the graph in Figure 3.4.

Proposition 3.6 *The fundamental group of the single-edge single-vertex directed graph (Figure 3.4) is isomorphic to* $(\mathbb{Z}, +)$.

Proof There is a natural isomorphism in which $a^n \cdot a^m = a^{n+m} \mapsto n + m$ for all $m, n \in \mathbb{Z}$. □

The torus has two homotopy classes of loop, as seen in Figure 3.5. Label the loops a and b. There is a homotopy such that $a \cdot b = b \cdot a$, since the combined loop can be continuously deformed (though temporarily breaking the subloops) along the surface of the torus in order to swap the elements without moving the base point.

Aside from this commutativity, the fundamental group of the 1-torus is otherwise free, hence,

Proposition 3.7 *The fundamental group of a 1-torus is isomorphic to the direct product* $\mathbb{Z} \times \mathbb{Z}$ *with* $+$ *(free abelian group of rank 2).*

One might wonder from the examples so far if there is any topological space in which the fundamental group is not commutative.

Definition 3.8 *An* n-rose *is a graph of one vertex and n loops at that vertex. Each loop in a rose is called a* petal.

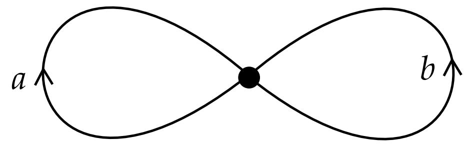

Figure 3.6: The figure-eight graph, with fundamental group $F_{\{a,b\}}$.

Figure 3.4 depicts a 1-rose, and Figure 3.6 shows a 2-rose. However, the 2-rose does *not* have the same fundamental group as that of the 1-torus. In the case of a torus, the surface provides homotopy such that loops are commutative. But it is tempting to suggest "$a \cdot b \psi b \cdot a$" in the 2-rose case by exhibiting the following function:

$$U(t,x) = \begin{cases} f_{a \cdot b}(x + \frac{t}{2}) & \text{if } x + \frac{t}{2} \leq 1 \\ f_{a \cdot b}(x + \frac{t}{2} - 1) & \text{if } x + \frac{t}{2} > 1 \end{cases}$$

What exactly is U doing here? It appears to continuously deform $a \cdot b$ into $b \cdot a$. Consider the position of the base point $f_{a \cdot b}(0)$. U attempts to move this point around the first half of the combined loop $a \cdot b$, $U(t, 0)$ is not constant. This is ruled out because it shifts the base point, so the fundamental group is not commutative.

Proposition 3.9 *The fundamental group of an n-rose is a free group of rank n ($n \geq 1$).*

Proof The case for $n = 1$ is given by Proposition 3.6. The rest is by induction. Suppose the result holds for a k-rose. A $(k + 1)$-rose is then the k-rose glued together with a 1-rose at the vertex. There is no homotopy that can make the 1-rose loop commute with the others without shifting the base point, so the fundamental group is, up to isomorphism, the free product $F_k * F_1 = F_{k+1}$. □

Note that the initial choice of base point does not affect any of these results, i.e. any base point could be used to find $\pi_1(X)$. The requirement is just that it remains fixed while considering homotopy classes.

3.5 Applications for fundamental groups

All graphs have free fundamental groups, and this characterises free groups (Wilton 2009).

A classical property of graphs (and surfaces also) is the *Euler characteristic*, denoted $\chi(G)$. For a graph $G = (V, E)$, $\chi(G)$ is given by the formula $\chi(G) = |V| - |E|$ (Post 2008). For a surface X, it is defined in terms of any cell decomposition Δ of X in the traditional way, that is, $\chi(X) = \chi(\Delta) = v - e + f$, the number of vertices v minus number of edges e plus number of cells f, the same formula for polyhedra. This can be thought of as the number of objects of dimension 0 (vertices) minus the number of objects of dimension 1 (edges) and so on. For abstracted spaces composed from objects of higher dimensions, it may be generalised by continuing this alternating sum.

Proposition 3.10 *A group is free if and only if it is the fundamental group of a graph $G = (V, E)$. In the finite case, the rank of the group is $1 - \chi(G)$.*

Proof It is already known that a free group of rank n is the fundamental group of an n-rose, thus the converse is sought.

Let $G = (V, E)$ be a graph, and T be a spanning tree of G (a subgraph without cycles including all vertices, and therefore having $|V| - 1$ edges). Collapse T to a single vertex—this is done by taking the quotient space G/T. G/T has $|V| - 1$ fewer vertices and $|V| - 1$ fewer edges than G, hence the quantity $|V| - |E|$ remains unchanged, but also the number of loops n is the same. (Note that the choice of T does not matter.) Thus G is homotopy equivalent to an n-rose, which has 1 vertex and n edges, so $|V| - |E| = 1 - n$, and by Proposition 3.9 has free group of rank n. Hence $n = 1 - \chi(G)$. □

The use of this characterisation helps with a suave proof of the following. The subgroups of a free group we feel must be free as well, since combinations of elements taken from the larger group, as a generating set, can only cancel in the same limited way. With some help from a correspondence between covering spaces and subgroups observed by Reidermeister, a totally different proof comes about (Stillwell 1993). The proof given by Schreier (1927) is in essence an algebraic transcription.

The index of a subgroup H of a group, denoted $|G : H|$, is the number of distinct left cosets[3] of H in G.

Theorem 3.11 (Neilsen-Schreier Theorem and the Schreier index formula) *Every subgroup of a free group is free. For a free group G, the rank of a subgroup H is given by* $|G : H|(\text{rank}(G) - 1) + 1$.

Proof By Proposition 2.9, G has simply transitive action consisting of graph automorphisms on its Cayley graph, a tree T. (T is here equivalent to the *covering graph* for the n-rose (see Fulton 1995)) Taking a subgroup H corresponds to a smaller covering graph T', also a tree, and since the subgroup will have the same kind of action on T', H must be free by Proposition 2.9.

Let $G \cong \pi_1(X)$ and $H \cong \pi_1(X')$. We use the fact that $\chi(X') = |G : H|\chi(X)$. By 3.10,

$$\begin{aligned}\text{rank}(H) &= 1 - \chi(X') \\ &= 1 - |G : H|\chi(X) \\ &= 1 - |G : H|(1 - \text{rank}(G)),\end{aligned}$$

and hence $\text{rank}(H) = |G : H|(\text{rank}(G) - 1) + 1$. □

By rewriting the index formula as

$$|G : H| = \frac{\text{rank}(H) - 1}{\text{rank}(G) - 1}$$

and noting that $|G : H|$ is always an integer, certain values for rank(H) can be ruled out for a given rank(G). (Stillwell 1993)

3.6 Presentations

F_2 has the *presentation* $\langle a, b|\emptyset\rangle$. This is thought of as having all possible words made from a, b, a^{-1}, b^{-1} with reduction as in the definition of a free group, and no *other* limitations, so the notation $\emptyset$ is used. To take another example,

[3] Alternatively, *right* cosets. The two formulations are consistent for any group and subgroup.

a free abelian group of rank 2 has presentation $\langle a, b|ab = ba\rangle$. There is now the limitation that only words that are reduced not only in the free group sense, but in addition, the commutative sense, may be allowed.

By the existence of identities and inverses, the relations of the form $l = r$ may be rewritten into the form $lr^{-1} = 1$ (equivalently, $l^{-1}r = 1$), where 1 is the identity of the group, and then the "= 1" part may be assumed. Thus the free abelian group has presentation $\left\langle a, b \middle| aba^{-1}b^{-1}\right\rangle$, where the relator $aba^{-1}b^{-1}$ is implied to be equal to the identity. It is nice to have another set of symbols to represent groups that we have already, but defining presentations takes advantage of free groups.

Definition 3.12 *A* group presentation *(or simply* presentation*)* $\langle S|R\rangle$ *is a set of* generators S *together with set of* relators $R \subseteq F_S$ *and is defined to be* F_S/N*, where* N *is the normal closure of* R *inside* F_S *(i.e. the smallest normal subgroup containing* R*).* $\langle S|R\rangle$ *is a presentation of a group* G *if* $G \cong \langle S|R\rangle$*.*

Recall the Neilsen-Schreier Theorem (3.11): every subgroup of a free group is free. This implies that the normal closure N of R is free. One might feel that if N is free, taking F_S/N will "chop out" generators completely, but this is not the case. For example, $\langle a|a^n\rangle$ is a presentation for $\mathbb{Z}_n$. a^n is a perfectly valid generator for a free group of rank 1, and the quotient $F_{\{a\}}/F_{\{a^n\}}$ is isomorphic to $\mathbb{Z}_n$, in no small part due to $F_{\{a\}} \cong F_{\{a^n\}}$.

When listing elements of S and R inside the angle brackets, it is convention to omit the braces $\{\cdot\}$, since the angle brackets do much the same job.

Every group has a presentation, even if an uncountably infinite number of generators and relators are required. The proof of this uses the universal property of free groups, which states that homomorphisms from $F_S \to G$ are in surjective correspondence with functions from $S \to G$.

Proposition 3.13 *Every group* G *has a presentation.*

Proof G generates itself. By the universal property of free groups (see Wilton 2009), it is possible to obtain G as a quotient $F_G/\mathrm{Ker}(\phi)$ for some ϕ being the homomorphism $\phi : F_G \to G$ corresponding to the identity map $I : G \to G$. $\square$

If there is a presentation for a group where both S and R are finite, the group is said to be *finitely presented*, likewise, if S is finite then it is *finitely generated*, and if R is finite then it is *finitely related*.

Products can now be defined succinctly as follows. If G has presentation $\langle S|R\rangle$ and H has presentation $\langle T|Q\rangle$, and S and T are disjoint, then the free product $G * H$ has presentation $\langle S \cup T|R \cup Q\rangle$. Likewise, the direct product $G \times H$ has presentation $\langle S \cup T|R \cup Q \cup [S, T]\rangle$, where $[S, T]$ is the set of all the commutators of elements of S with elements of T.

3.7 Diagrams

The motivation behind diagrams is to *geometrically* discover the consequences of relators in a presentation. The idea is to create a diagram, consisting of an oriented cell decomposition with labels on the edges consistent with a presentation, and reading off consequences.

We can deduce consequences symbolically. A simple example is to deduce that $c^4 = 1$ from the presentation $\left\langle a, b, c \middle| a^4, bab^{-1}c \right\rangle$. Now suppose that we do not know in advance that $c^4 = 1$. It is only a few deductive steps away from the given relators, but it might be the case that such steps are never taken. For more complicated groups, to discover consequences in this manner can be tedious, or even impossible. Diagrams provide an alternate means for discovering consequences, although are largely subject to the same limitations. The process for creating an appropriate diagram is as follows.

Take a cell decomposition Δ of a surface X. In this context, Δ is called a *map on* X. If X is orientable, choose some orientation for Δ. Depending on X, Δ may be referred to as planar, spherical, toroidal, annular, an so on, in the same manner as embedded graphs. Now, if the surface has a boundary, then Δ will have some component C consisting of edges $e_1, e_2, \ldots, e_n$ sitting on $\partial\Delta$, connected head-to-tail, and furthermore, they will form a loop. Each such loop is called a *contour* of the map Δ. Contours of a circular map are comparable to the boundary of an individual cell, and for a circular map $\partial\Delta$ coincides with the contour.

Contours are equivalent up to changing the starting edge, an equivalence termed *up to cyclic shift*. Specifically, the contour $e_1 e_2 \dots e_n$ is equivalent to $e_k e_{k+1} \dots e_{n-1} e_n e_1 e_2 \dots e_{k-1}$ for every $1 \leq k \leq n$. When either e or e^{-1} (the reverse edge of e) is one of the edges in a contour, that edge is said to *belong to the contour*.

Note that the number of distinct contours of a map varies with the surface X, and we are interested in using surfaces with at least one contour when creating a diagram.

To make the map useful, the edges require labelling. In the same vein as Definition 2.1, symbols from an alphabet S each have some (tentatively-named) inverse symbol a^{-1} in S^{-1}, and there is an identity symbol (henceforth 1, with $1^{-1} = 1$) also adjoined as necessary, in order to obtain the set of possible labels S^1 defined as $S^1 = S \cup S^{-1} \cup \{1\}$.

Definition 3.14 *For a given alphabet S, a map having a set of edges E is a* diagram over S *when there is an edge labelling $\phi : E \to S^1$ such that for all $e \in E$, $\phi(e^{-1}) = \phi(e)^{-1}$, i.e. reverse edges have the inverse label.*

To perform "reading" of a diagram, take the strings of labels along paths. This naturally extends the labelling ϕ to include paths. For each path $p = e_1 e_2 \dots e_n$, then define $\phi(p) = \phi(e_1)\phi(e_2)\dots\phi(e_n)$, and for an empty path (consisting of $n = 0$ edges) define $\phi(p) = 1$. As for a contour C, the labelling ϕ should also be defined up to taking cyclic shifts.

The next trick is relating this to a presentation $\langle S|R \rangle$. Consider a cell Π, having a cell contour p. If $\phi(p)$ is *visually equal* to some relator W in R, Π is an *R-cell*. Of course we might accidentally read the cell contour the wrong way around, or perhaps have a few 1 edges in there, so if $\phi(p)$ is visually equal—up to cyclic shift—to W, W^{-1}, or either W or W^{-1} with any number of 1 symbols inserted, then it is an R-cell. Figure 3.7 shows a diagram over $\{a, b, c\}^1$ where each cell is an R-cell in $\left\langle a, b, c \middle| a^4, b^{-1}abc \right\rangle$.

The other type of cell that occurs in diagrams of interest are *0-cells*. 0-cells either have only *0-edges* (edges with label 1), or there is one pair of letters (not necessarily consecutive edges) labelled from S^1 that cancel with each other (some a followed by a^{-1}), the remainder being 0-edges. So finally,

Definition 3.15 *For a given presentation $\langle S|R \rangle$, a diagram over S is a* diagram over $\langle S|R \rangle$ *(or diagram over G if $G \cong \langle S|R \rangle$) when all cells are either R-cells or 0-cells.*

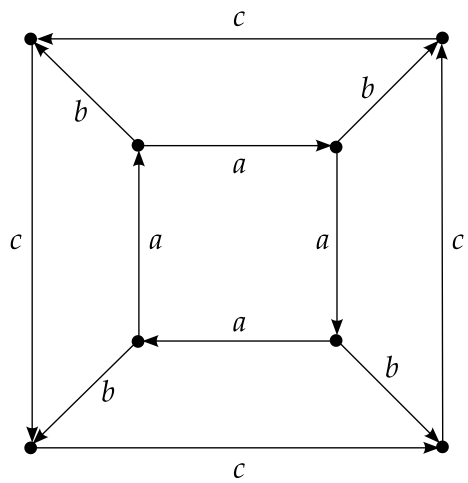

Figure 3.7: An example diagram over $\left\langle a, b, c \middle| a^4, b^{-1}abc \right\rangle$, from which $c^4 = 1$ can be deduced from reading the label of the outer contour.

The similarity between diagrams and Cayley graphs seems strong. They both encode group structure into a graph. The main differences are that diagrams include edges labelled with the identity, where Cayley graphs do not, that a Cayley graph includes every group element represented by vertices whereas a diagram does not associate any particular label to any vertex and there isn't a guaranteed number of them. Also, diagrams are explicitly cell decompositions, whereas it is not obvious at all how to embed any given Cayley graph into a surface to form a cell decomposition—furthermore, some Cayley graphs, such as the Cayley graph of a free group, are collapsible to a single point anyway.

The following result is key to the usefulness of a diagram.

Lemma 3.16 (van Kempen's Lemma) *Let w be an arbitrary non-empty word in S^{1*}. Under the presentation $\langle S|R\rangle$, $w = 1$ if and only if there exists a circular diagram over $\langle S|R\rangle$ with contour label w.*

Such a circular diagram is called the *diagram of deduction* of the consequence $w = 1$ over $\langle S|R\rangle$. Given an existing diagram of deduction, it is often possible to add additional R or 0-cells *ad infinitum* to obtain more diagrams. This parallels using relators as steps in a symbolic deduction, but provides another means for doing so. We can go further, however.

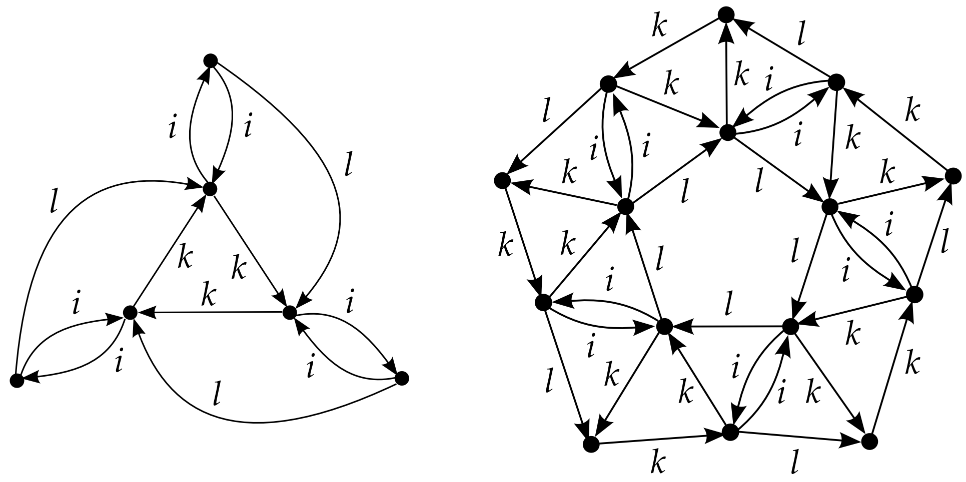

Figure 3.8: Diagrams over the icosahedral group having presentation $\left\langle i,k,l \middle| i^2,k^3,l^5,ikl^{-1}\right\rangle$. Left: diagram of deduction for $(li)^3 = 1$. Right: diagram of deduction for $(kl)^5 = 1$, and also that l^5 and $(kl)^5$ are conjugates.

Lemma 3.17 *Let v, w be non-empty words in S^{1*}. Then under the presentation $\langle S|R\rangle$, v and w are conjugate (i.e. $w = ava^{-1}$ for some $a \in \langle S|R\rangle$) if and only if there is an annular diagram over $\langle S|R\rangle$ with (distinct) contours p, q such that the label of p is v and the label of q is w.*

There are many simple operations that can be applied to a diagram that leave it as a valid diagram, such as most ways of inserting or removing R- or 0-cells.

Some more examples of diagrams are in Figure 3.8, this time over the icosahedral group. One of the first known examples of a group presentation, though it was yet to be called that, was Hamilton's "Icosian Calculus," nowadays known as the icosahedral group. The presentation given by Hamilton (1856) was[4] $\left\langle \iota,\kappa,\lambda \middle| \iota^2 = 1, \kappa^3 = 1, \lambda^5 = 1, \iota\kappa = \lambda\right\rangle$, but this is isomorphic (by $i \equiv \iota, k \equiv \kappa, ik \equiv \lambda$) to the conciser $\left\langle i,k \middle| i^2,k^3,(ik)^5\right\rangle$.

[4]Although as the concept of presentation hadn't yet been developed, Hamilton didn't use angle brackets.

3.8 Small cancellation theory

The utility of van Kempen's lemma was inadequately realised for almost 30 years following its publication in 1933. During the 1960s, small cancellation theory was invented and van Kempen's lemma finally received some overdue attention, as diagrams were of central importance. The core idea is that seemingly abstract conditions on a group can affect the geometry of possible diagrams in useful ways, and therefore the possible deductions.

Cyclic shifts and taking inverses might again cause inconvenience when talking about relators, so as a matter of convenience, we take a set of relators R to be *symmetrised*, meaning that for any $r \in R$ we require r^{-1} and cyclic shifts of r and r^{-1} to be in R. In describing cancellation conditions the relators are assumed to be symmetrised.

Now suppose $r_1, r_2 \in R$ for a presentation $\langle S|R\rangle$, and that they start (on the left) with the same substring x, so $r_1 = xy_1$ and $r_2 = xy_2$ (or such a form for r_1, r_2 as obtainable by cyclic shifts and inverses). x is then known as a *piece*.

Immediately, for $|w|$ denoting the length of the string w, we have

Definition 3.18 *For a given $\lambda \in [0,1]$, a group $G \cong \langle S|R\rangle$ satisfies* the condition $C'(\lambda)$ *if $|x| < \lambda|r|$ for all $r \in R$ and pieces x.*

Suppose G satisfies $C'(\lambda)$, and let $r_1 = xy_1, r_2 = xy_2$ be relators sharing the piece x. One way of obtaining a consequence in G is to now take $r_1^{-1}r_2 = y_1^{-1}x^{-1}xy_2 = y_1^{-1}y_2$—the piece x is that much of the relators that can cancel. The condition $C'(\lambda)$ now specifies that the amount of any such cancellation amongst relators is going to be bounded above by the factor λ of the word length of the relators. Hence the term *small cancellation*.

There is another cancellation condition, denoted $C(k)$, meaning for a presentation $\langle S|R\rangle$ that every $r \in R$ must be composed of at least k pieces.

Small cancellation conditions interact with diagrams in a neat way. Suppose we form some cells corresponding to relators. The stricter the condition $C'(\lambda)$, i.e. the smaller the λ, the more of each cell is unique to the relator, and the less places (corresponding to pieces) there will be to glue them together.

3.8.1 Uses for small cancellation theory

Small cancellation conditions, and van Kempen diagrams, provide the foundation for many results in geometric group theory over the last 30 years, especially the formation of the Tarski monster and other related counterexamples to the von Neumann conjecture (Ol'shanskii and Sapir 2002).

Another application is in showing Dehn's algorithm, for solving the word problem, works on all finitely presented groups satisfying $C'(\lambda)$ for $\lambda \leq \frac{1}{6}$.

Another application is that $C(6)$ presentations are *aspherical* (that a group does not admit non-trivial reduced graded spherical diagrams). The free abelian group of rank 2 is an example of an aspherical group, but the free abelian group of rank 3 is not aspherical. Asphericity, in turn, implies that all the relators in a presentation are independent of one another, that is, no relator can be deduced from any of the others. To show asphericity, one can use inequalities based on the relationship between relators and R-cells to contradict $\chi(\Delta) = 2$ for each possible spherical cell decomposition Δ. Finally, asphericity is the subject of the Whitehead conjecture, which states every subpresentation of an aspherical presentation is aspherical, and has been the subject of recent attention (Ivanov 1998).

Chapter 4

Geometry's Revenge

In section 2.5, it is remarked that paradoxicality is a property of equidecomposability classes, and since paradoxicality implies there are at least two disjoint subsets equidecomposable to a set, each of which therefore inherit paradoxicality, the true nature of paradoxicality is as a Cantor set-like structure on descending subsets all equidecomposable to one another. So, the non-measurable pieces used to duplicate a sphere can be thought of as a combination of fractal self-similarity and the Axiom of Choice. While the Axiom of Choice adequately prevents describing the pieces constructively, it is still of interest to flesh out the contribution of fractal self-similarity. This chapter looks at how F_2 so conveniently provides this in its own structure, with a particular view to its own geometry and related topics.

Recall that F_2 has the group presentation $\langle a, b | \emptyset \rangle$. The Cayley graph of F_2 is shown again in Figure 4.1.

The graph as illustrated in Figure 4.1 exhibits self-similarity. This occurs partially because it must fit in a finite area of paper. If the Euclidean length of any given path on the diagram were proportional to the word metric, matters would be significantly more difficult. To circumvent this problem here, the path from the identity to any particular element is made to have length that converges exponentially to some finite length as the word length becomes large.

A related question: in what metrizable surfaces can the Cayley graph of F_2 (or, for generality, any group) be embedded, such that the graph topology is fully compatible with that of the surface?

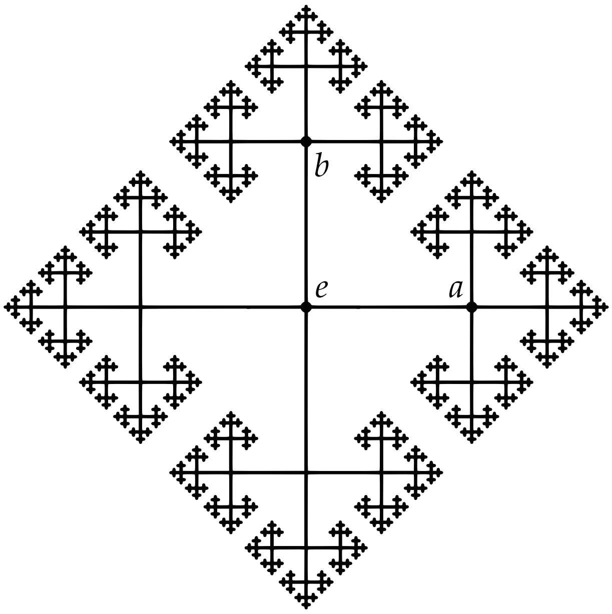

Figure 4.1: The Cayley graph for F_2, with the generators a, b and identity e labelled.

4.1 L-systems

The illustration in Figure 4.1 can be more formally described with an L-system.

Definition 4.1 *A deterministic* L-system *is a tuple* (S, A, Φ) *consisting of an alphabet* S*, a string* $A \in S^*$ *called the* axiom*, and a set of* substitution rules $\Phi \subseteq (S^*, S^*)$*, which provides for each of some substrings another substring to replace it. Let* $V : S^* \to S^*$ *denote the operation that performs as many substitutions described in* Φ *as possible simultaneously. An L-system defines at least countably many* generations*, each a string from* S^**, where generation* n *is given by* $V^n(A)$*, that is, the generations* $0, 1, 2, \ldots$ *are* $A, V(A), V(V(A)), \ldots$.

The operation V is successively iterated, starting with the axiom A, so L-systems are examples of iterated function systems. L-systems and iterated function systems are popular for producing fractal and fractal-like figures (where each letter describes a drawing action) and simulating lifeforms (each letter representing, for instance, an artificial bacterium), and other offensive purposes, such as the virtual road-network generation of Nobuko et al. (2000).

Figure 4.1 was drawn using an L-system with a formulation similar to the following:

Let r, u, r^{-1}, u^{-1} represent the movement of a "pen" upon "paper"[1] in the directions right, up, left, and down, respectively, and let $+, -$ denote increasing and decreasing the subsequent lengths of movement by a factor of 2. Then, the Cayley graph for F_2 can be drawn within a finite area to any desired finite depth (though only a few generations can be enough to communicate the point) starting with the axiom of $rr^{-1}uu^{-1}r^{-1}ru^{-1}u$ and using the substitution rules

$$\begin{aligned} rr^{-1} &\to r{-}u^{-1}urr^{-1}uu^{-1}{+}r^{-1} \\ r^{-1}r &\to r^{-1}{-}u^{-1}ur^{-1}ruu^{-1}{+}r \\ uu^{-1} &\to u{-}r^{-1}ruu^{-1}rr^{-1}{+}u^{-1} \\ u^{-1}u &\to u^{-1}{-}r^{-1}ru^{-1}urr^{-1}{+}u. \end{aligned}$$

Suppose the string at any generation has the operator R (of Definition 2.1) applied to it. If $+$ and $-$ are considered inverses for the purposes of cancellation, the entire string will vanish at each generation.

4.2 Box-counting

Briefly, a *fractal dimension* is a measure of how complicated a fractal is, designed to handle objects that lie seemingly between integer dimensions. An easy example to comprehend is the box-counting dimension.

Definition 4.2 *Let A be a bounded subset of the plane. Consider an equi-spaced square grid with squares of side ϵ superimposed upon A. Let $N(\epsilon)$ be the number of squares from the grid having interior which is not disjoint from A. The* box-counting dimension *of A is then defined as*

$$\dim_{\text{box}}(A) = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}}$$

where such a limit exists.

[1] Computer simulations are acceptable.

The box-counting dimension is related to the Hausdorff dimension, packing dimension, and many other similar measures of fractal dimension (not defined here) involving similar ideas. In exotic cases where the limit does not exist, the upper and lower Minkowski dimensions, corresponding *mutatis mutandis* to limit superior and limit inferior for limit in Definition 4.2, may still be defined.

Consider one of the four sub-3-trees of the F_2-Cayley graph-L-system-drawing of Figure 4.1. Let $N(n)$ be the number of squares of side $\epsilon = 2^{-n}$ required to cover one of these, and suppose for convenience that a distance of 1 happens to be the side of a bounding square of the whole tree, so $N(0) = 1$.

Figure 4.2 depicts the first few steps of one attempt at this, where the grid is in the same direction as the tree. It is possible, although cumbersome, to obtain a formula for N from this particular choice of alignment. To see why, consider a recurrence on the 3 similar subtrees. $N(n-1)$ will be counted 3 times in $N(n)$, but there are some overlapping squares, for instance, the square covering the bottom tip of the "trunks" of the subtrees will be counted 3 times, and furthermore some of the squares covering the "foliage" also overlap. Nevertheless, it is clear that the driving term will be some multiple of 3^n.

The alignment in Figure 4.3, on the other hand, is far easier to work with. Each of the 3 subtrees are boxed precisely as in the previous tree with no overlaps between them, so the only remaining step is to cover the trunk, requiring 2^{n-1} squares. Thus this recursive formulation:

$$N(0) = 1, \quad N(n) = 3N(n-1) + 2^{n-1}$$

which takes the integer sequence $1, 4, 14, 46, 146, 454, \ldots$. Sloane (2009a) gives the closed form $N(n) = 2 \cdot 3^n - 2^n$ (in addition to the recursive one above).

Hence

$$\begin{aligned} \dim_{\text{box}} &= \lim_{n\to\infty} \frac{\log N(n)}{\log 2^n} \\ &= \lim_{n\to\infty} \frac{\log(2 \cdot 3^n - 2^n)}{\log 2^n} \\ &= \frac{\log 3}{\log 2} \approx 1.585 \end{aligned}$$

Since this formulation permits four copies of the tree to be attached without overlaps as in the Cayley graph of F_2, this is also the box-counting dimension for the entire graph as illustrated by the L-system.

Is it possible, for any given integer sequence, to find a fractal with the box-count $N(n)$ taking the values from that sequence? If attention is restricted to L-systems as described so far, which are computable, the sequence must be at least computable.

What about applying box-counting, or a similar idea, to the points formed on a sphere, as a result of taking an orbit of the free subgroup of SO_3, for example, generated by two rotations such as the A, B of Theorem 2.10?

While box-counting generalises easily to higher dimensions, the use of boxes might be difficult, since no side of a box lines up very well with the curved surface. Box-counting is an *extrinsic* dimension, where the unit of measurement is imposed via the space in which the set being measured is embedded. *Intrinsic* dimensions do not depend on the space, rather, only subsets of the set in question are used as the unit of measurement. In the case of Figure 4.1, balls of a given radius contained in the graph topology are easily defined. However, applying intrinsic dimensions to the case of an orbit of $F_{\{A,B\}}$ on a point from S^2 appears to be equally difficult as for extrinsic dimensions. It is not even clear whether or not the orbit is dense in S^2.

4.3 L-systems and expressions

While the L-system above gives a nice pictorial representation of F_2, it is possible to use an L-system-like construction to explicitly list all the elements in F_2 of length n at the nth generation. Consider the axiom $\{e\} \cup S(a) \cup S(b) \cup S(a^{-1}) \cup S(b^{-1})$ and the substitution rule *schemas* (where, on the left, σ matches the remainder of the string inside the containing parentheses):

$$\begin{aligned}
S(\sigma a) &\to \{\sigma a\} \cup S(\sigma aa) \cup S(\sigma ab) \cup S(\sigma ab^{-1}) \\
S(\sigma b) &\to \{\sigma b\} \cup S(\sigma ba) \cup S(\sigma bb) \cup S(\sigma ba^{-1}) \\
S(\sigma a^{-1}) &\to \left\{\sigma a^{-1}\right\} \cup S(\sigma a^{-1}b) \cup S(\sigma a^{-1}a^{-1}) \cup S(\sigma a^{-1}b^{-1}) \\
S(\sigma b^{-1}) &\to \left\{\sigma b^{-1}\right\} \cup S(\sigma b^{-1}a) \cup S(\sigma b^{-1}a^{-1}) \cup S(\sigma b^{-1}b^{-1})
\end{aligned}$$

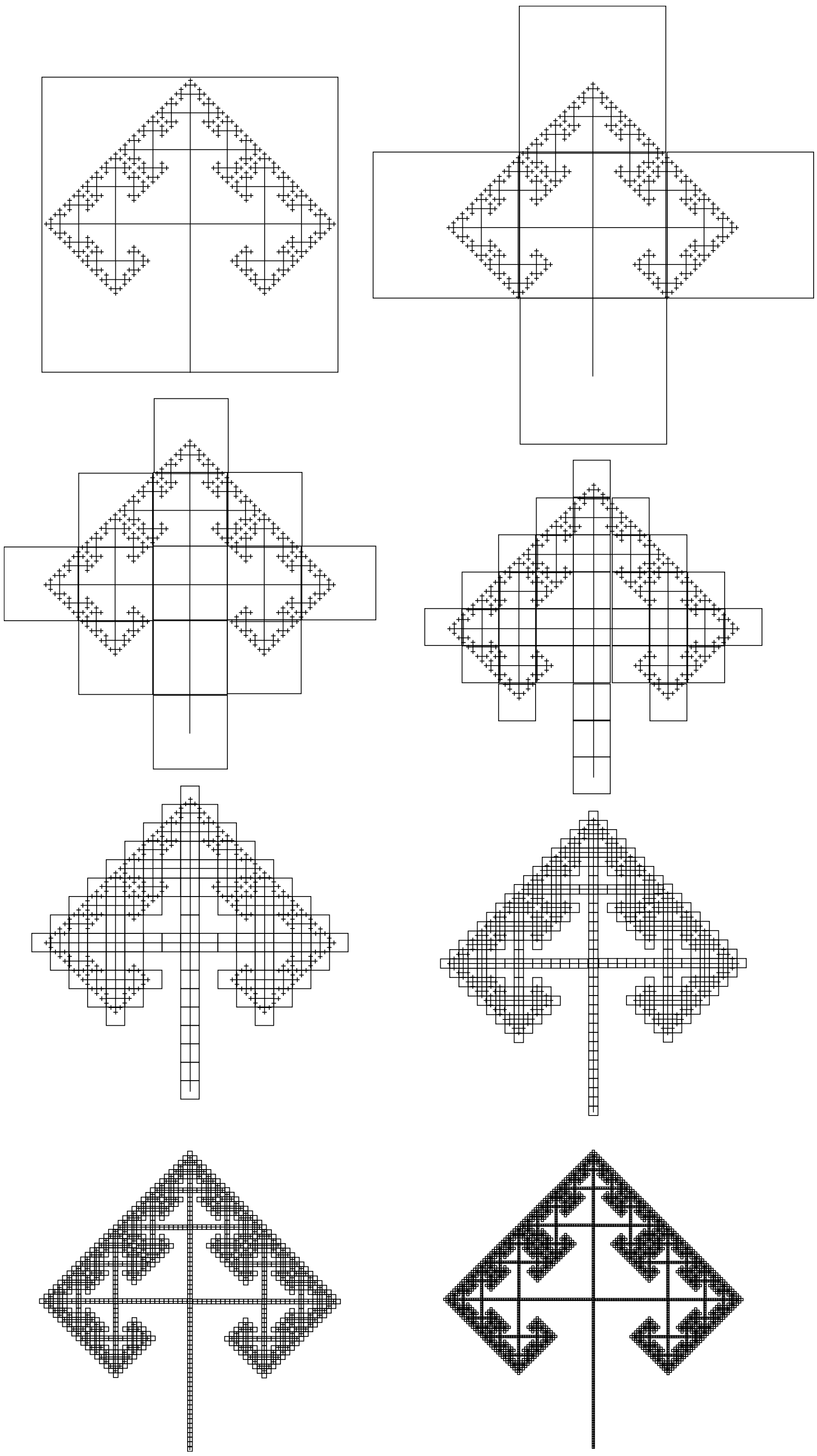

Figure 4.2: A basic demonstration of box-counting the 3-tree fractal, for $n = 0, 1 \ldots, 7$.

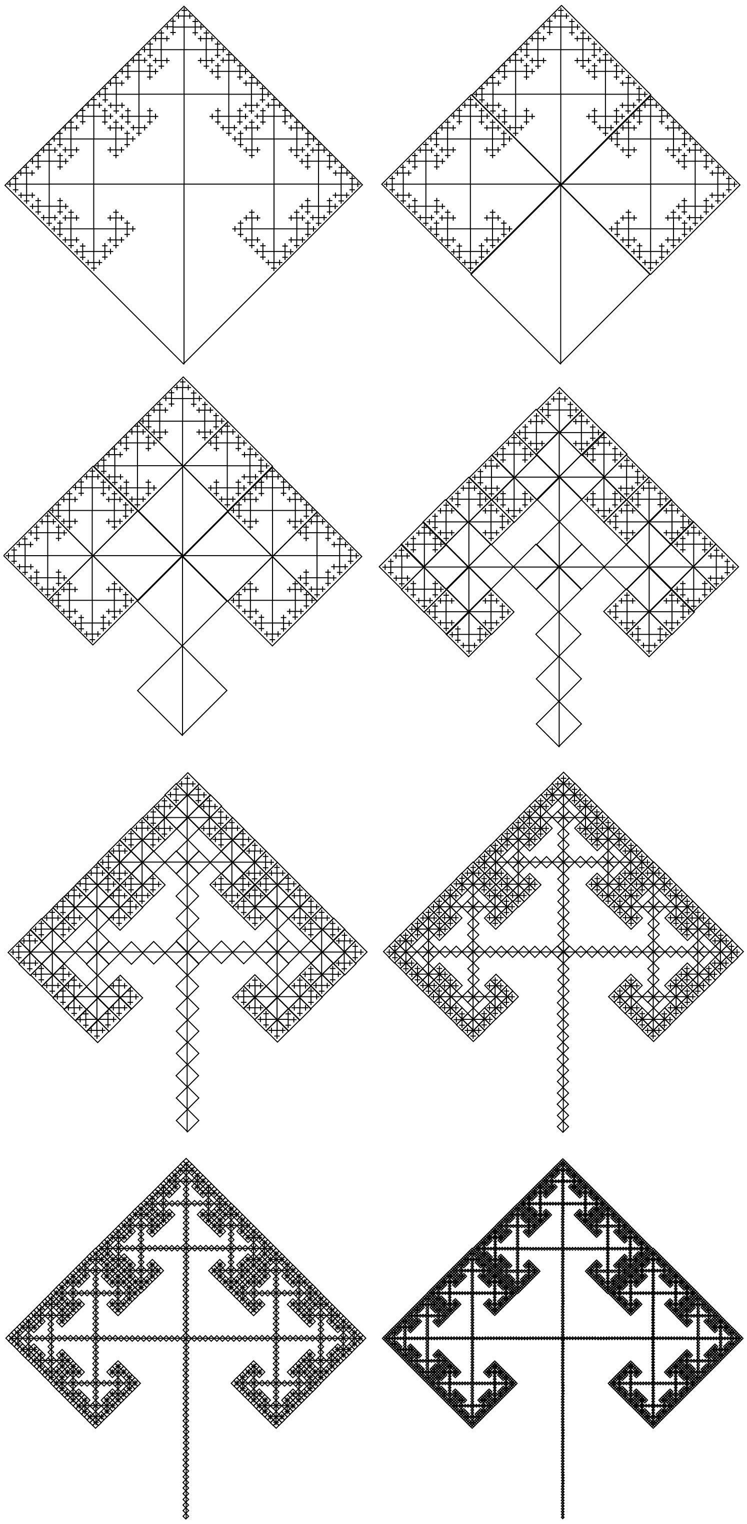

Figure 4.3: A more effective alignment for box-counting the 3-tree for $n = 0, 1 \ldots, 7$.

If S is assigned its usual meaning (that is, $S(\sigma)$ = set of reduced strings starting with σ), then every generation will be the set F_2. But suppose instead that S does not—suppose $S(\sigma) = \emptyset \forall \sigma$. Now the nth generation corresponds to

$$\{\sigma \in F_2 : |\sigma| \leq n\}$$

where $|\cdot|$ is the word length. Call this set $F_2[0,n]$. Immediately, $F_2 = \bigcup_{n=1}^{\infty} F_2[0,n]$ and even $F_2 = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} F_2[0,n]$, that is, $F_2 = \limsup_{n\to\infty} F_2[0,n]$. Furthermore, each $w \in F_2$ appears in all but finitely many $F_2[0,n]$, so $F_2 = \liminf_{n\to\infty} F_2[0,n]$.

Also,

$$F_2[0,n] \backslash F_2[0,n-1] = \{\sigma \in F_2 : |\sigma| = n\}\,.$$

so call this one $F_2(n)$. This is the number of "leaves" in the nth generation of the Cayley graph of F_2. Now

$$|F_2(n)| = 3|F_2(n-1)| \quad \text{for } n \geq 2$$

The terms in this sequence are $1, 5, 17, 53, \ldots$. Each step adds 3 times the number of "leaves" from the previous, and the number of leaves proceeds in the sequence $0, 4, 4\cdot 3, 4\cdot 3^2, \ldots$. Thus the sequence of values taken by $F_2[0,n]$ satisfies the recurrence relation

$$\begin{aligned} |F_2[0,n]| &= 3|F_2(n-1)| + |F_2[0,n-1]| \\ &= 3|F_2[0,n-1] \backslash F_2[0,n-2]| + |F_2[0,n-1]| \end{aligned}$$

or, in closed form, $|F_2[0,n]| = 2\cdot 3^n - 1$ (Sloane 2009b). An interesting property of this sequence is that it is also the number of triangles (including "holes") in the Sierpinski gasket fractal after n inscriptions.

The growth property of a group is important in the study of infinite groups. The above sequence shows the growth of the free group to be exponential.

There are three classes of growth: exponential, polynomial, and intermediate (of subexponential growth but greater than polynomial growth). Groups of intermediate growth were conjectured to exist by Milnor in 1968 and remained in that state as recently as 1984, at which point Grigorchuk introduced some examples (Grigorchuk and Pak 2006).

Note that both the sequence $N(n)$ of boxes of side $\frac{1}{2^n}$ required to cover the tree fractal is given by a driving term of 3^n, and the growth of the group is order of 3^n.

4.4 Geometry of Cayley graphs

Suppose a Cayley graph for some group G is endowed with the following metric: $d(\sigma, \tau)$ is the number of edges in the shortest path between the vertices for σ and τ in the Cayley graph. Each edge is homeomorphic to the unit interval, so we also define d for points between vertices.

For F_2, the closed ball centred at the identity vertex e of radius n has all the vertices labelled with members of $F_2(n)$, that is, $\overline{B_n}(e)$ contains the part of the graph corresponding to $F_2(n)$. Again, growth of a group could be defined in terms of the sequence taken by $|B_n(e)|$.

Metric spaces are capable of more than simply defining balls. Consider a shortest path along edges in a Cayley graph. Such a path is called *geodesic*. A triangle in the Cayley graph, consisting of three vertices and the paths connecting them, is geodesic if the paths are.

Small cancellation theory imposed a condition on the symbolic facet of a group, that is, the composition of relators in terms of subwords, which had a visible affect on the geometric facet of groups as in van Kempen diagrams. Can a condition be imposed in the goemetric facet that goes the other way?

Yes. If each side of a geodesic triangle is contained in a ball of some fixed positive radius δ containing each of the other two sides, then the triangle is *δ-thin*. Now suppose all possible triangles in a Cayley graph are δ-thin—if this is the case, the Cayley graph (and group) are *δ-hyperbolic* (or simply *hyperbolic*). Hyperbolic groups were introduced by Gromov, who also investigated the connection between growth conditions and metric spaces on groups (for example, Gromov 1981). This terminology is of course inspired from more general metric spaces of non-positive curvature (Bridson and Haefliger 1999). There is a hyperbolic surface in which the Cayley graph of a hyperbolic group can be embedded such that the graph metric space and topology of the hyperbolic surface completely agree.

Unsurprisingly, the free groups generated by any finite set are examples of a hyperbolic groups. Just as $C'(\lambda)$ groups for $\lambda \leq \frac{1}{6}$ are solvable by Dehn's algorithm, it turns out that hyperbolic groups are also solvable. (Sankaran 2004)

4.5 L-systems and presentations

It is not clear whether L-systems can be systematically used, in general, to produce Cayley graphs, van Kempen diagrams, Schreier coset graphs or other useful geometric tools for studying arbitrary groups. But they are hardly useless.

A very modern development in the study of group theory has been the *endomorphic presentation*, or *L-presentation* (Bartholdi 2003).

Consider the problem of describing a non-finitely related group, that is, one requiring a presentation $\langle S|R\rangle$ where R is not finite. If one could identify a pattern among the relators, say in terms of recursively-applied substitution rules, then an L-system could be used to describe them all in a systematic fashion. Such is then an L-presentation.

Even in finitely-presented groups or finite groups, the use of L-presentations can make even more concise the symbolic description of a group. Since all groups admit a presentation, they also admit an L-presentation. With the introduction of L-presentations, lines of investigation into new and interesting properties of a group, such as finite L-presentability, have opened up.

Chapter 5

Conclusion

A core object used throughout, the proof of the Banach-Tarski paradox, and as an example in discussing homotopy groups, hyperbolic groups, and L-systems, is the free group. Each harnesses some idea or property related to the free group.

Free groups are, in a sense, a simple structure, having presentations without relators. The lack of relators maps directly to the uncomplicated topologies of which they are the fundamental groups. The 2-rose seems to be somewhat trivial. Despite this, free groups do not admit easy geometric interpretation as motions or transformations.

In another sense, free groups are complicated. F_2 is paradoxical under itself with pieces that are self-similar, and it has an intricate, non-obvious isomorphism to subgroups of SO_3. The Cayley graph of F_2 looks intricate. On the other hand, it is a covering graph of the 2-rose. If the 2-rose and presentation $\langle a, b|\emptyset\rangle$ illustrate the straightforward nature of F_2, the Cayley graph and self-similarity illustrate its complexity.

Even so, it is fortunate that the free group admits visualisation. What do other non-amenable groups look like? Answering this question was an initial goal of this thesis, though it soon became apparent that it was somewhat lofty. Ol'shanskii's first example was an infinite group having cyclic subgroups with the same prime order p, but with $p \geq 10^{75}$. It would appear large amounts of detail will be lacking from a depiction, since 10^{75} is many orders of magnitude larger than the number of known particles composing the pages of this thesis.

Another example of a non-amenable group is a group having finite presentation. The presentation, being the extension of a group with exponent n (large) by a cyclic group, satisfies $[x, y]^n = 1$ (Ol'shanskii and Sapir 2002). This may be easier to digest, but still leaves us with no straightforward way of producing a "big-picture" illustration of the group. How ironic that the proofs of non-amenability for these un-visual counterexamples to the von Neumann conjecture rely on geometric group theoretic techniques! In fact, the geometric group theory approaches to proofs suffer from a kind of un-reversability. We embed the structure of a presentation into a diagram, then use the geometry of the diagram to provide necessary conditions, and within these conditions we obtain the proof of a desired theorem, usually by contradiction. The proof is done in a world two steps away from the structure that we start with. However, the conditions on the geometry of a diagram over a group do not generally provide equivalent algebraic conditions, and this is where attempts to depict these complicated groups by reverse engineering fall over.

How is any of this different for F_2? With F_2, no geometric group theory of the above kind is required. Once again, the self-similarity of F_2 comes to the rescue. The hope in approaching F_2 from this angle was that similar self-similarity might be found in the more complicated examples of non-amenability, but this is far from clear.

Bibliography

Austen, J. and Grahame-Smith, S.: 2009, *Pride and Prejudice and Zombies*, Quirk Productions, Inc.

Banach, S. and Tarski, A.: 1924, Sur la décomposition des ensembles de points en parties respectivement crongruentes, *Fundamenta Mathematicae* **6**, 244–277.

Bartholdi, L.: 2003, Endomorphic presentations of branch groups, *Journal of Algebra* **268**(2), 419 – 443.

Boyer, J. M. and Myrvold, W. J.: 2004, On the Cutting Edge: Simplified $O(n)$ Planarity by Edge Addition, *Journal of Graph Algorithms and Applications* **8**(3), 241–273.

Bridson, M. R. and Haefliger, A.: 1999, Metric spaces of non-positive curvature, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **319**, xxii+643.

Foreman, M. and Wehrung, F.: 1991, The Hahn-Banach theorem implies the existance of a non-Lebesgue measurable set, *Fundamenta Mathematicae* **138**, 13–19.

Fulton, W.: 1995, *Algebraic Topology / A First Course*, Vol. 153 of *Graduate Texts in Mathematics*, Springer-Verlag.

Grigorchuk, R. and Pak, I.: 2006, Groups of intermediate growth: an introduction for beginners. Online; last checked 28 October 2009.
URL: `http://arxiv.org/abs/math.GR/0607384`

Gromov, M.: 1981, Groups of polynomial growth and expanding maps, *Publications Mathématiques de L'IHÉS* **53**(1), 53–78.

Hamilton, W. R.: 1856, Memorandum respecting a new system of roots of unity, *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science* **xii**, 446. Online; last checked 18 October 2009.
URL: `http://www.maths.tcd.ie/pub/HistMath/People/Hamilton/Icosian/NewSys.pdf`

Hausdorff, F.: 1914, Bemerkung über den Inhalt von Punktmengen, *Mathematische Annalen* **75**, 428–433.

Hazewinkel, M.: 2002, Hilbert infinite hotel, Springer Online Encyclopaedia of Mathematics, last checked 30 October 2009.
URL: `http://eom.springer.de/h/h130080.htm`

Hinkis, A.: n.d., Cantor's proof of the Cantor-Bernstein theorem, Online; last checked 30 October 2009.
URL: `http://www.hinkis.org/Mathematical_papers/01_03_Cantor's_CBT_proof_05.pdf`

Ivanov, S. V.: 1998, On aspherical presentations of groups, *Electronic Research Announcements of the American Mathematical Society* **4**, 109–114.

Kuratowski, C.: 1930, Sur le problème des courbe gauche en Topologie, *Fundamenta Mathematicae* **15**, 271–283.

Lakatos, I.: 1976, *Proofs and Refutations / The Logic of Mathematical Discovery*, Cambridge University Press.

Mulzer, W. and Rote, G.: 2008, Minimum-weight triangulation is NP-hard, *Journal of the ACM* **55**(2).

Mycielski, J. and Świerczkowski, S.: 1958, On free groups of motions and decompositions of the Euclidean space, *Fundamenta Mathematicae* **45**, 283–291.

Nobuko, K., Tomoe, O., Hitoshi, K. and Seiichi, N.: 2000, L-system Approach to Generating Road Networks for Virtual Cities, *Transactions of Information Processing Society of Japan* **41**(4), 1104–1112.

Ol'shanskii, A. Y.: 1991, *Geometry of Defining Relations in Groups*, Mathematics and Its Applications (Soviet Series), Kluwer Academic Publishers. Translated from the 1989 original by Yu. A. Bakhturin.

Ol'shanskii, A. Y. and Sapir, M. V.: 2002, Non-amenable finitely presented torsion-by-cyclic groups, *Publications Mathématiques de L'IHÉS* **96**.

Pawlikowski, J.: 1991, The Hahn-Banach theorem implies the Banach-Tarski paradox, *Fundamenta Mathematicae* **138**, 21–22.

Post, O.: 2008, Spectral analysis of metric graphs and related spaces, EPFL Press.
URL: `http://arxiv.org/abs/0712.1507`

Sankaran, P.: 2004, Lectures on hyperbolic groups. Online; last checked 28 October 2009.
URL: `http://www.imsc.res.in/~sankaran/hyper1.ps`

Schreier, O.: 1927, Die Untergruppen der freien Gruppen, *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* **5**(1), 161–183.

Sloane, N. J.: 2009a, A027649, Online Encyclopaedia of Integer Sequences. Online, last checked 18 October 2009.
URL: `http://www.research.att.com/~njas/sequences/A027649`

Sloane, N. J.: 2009b, A048473, Online Encyclopaedia of Integer Sequences. Online, last checked 28 October 2009.
URL: `http://www.research.att.com/~njas/sequences/A048473`

Stillwell, J.: 1993, *Classical topology and combinatorial group theory*, Vol. 72 of *Graduate Texts in Mathematics*, Springer-Verlag.

von Neumann, J.: 1929, Zur allgemeinen theorie des masses, *Fundamenta Mathematicae* **13**, 73–116.

Wagon, S.: 1985, *The Banach-Tarski Paradox*, Vol. 24 of *Encyclopedia of Mathematics and Its Applications*, Cambridge University Press.

Wilton, H.: 2009, Free groups - Notes from 392C Geometric Group Theory, Online, last checked 4 October 2009.
URL: `http://392c.wordpress.com/2009/01/21/1-free-groups/`