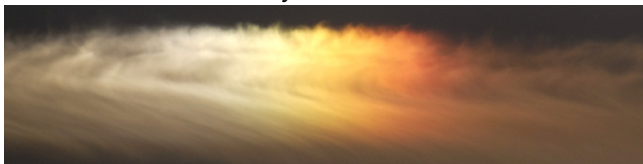


# Thompson groups, Cantor space, and foldings of de Bruijn graphs

Peter J. Cameron  
University of St Andrews



Breaking the boundaries  
University of Sussex  
April 2015

## The 1970s

I was born (in Paul Erdős' use of the word) in 1971, as a finite group theorist, specifically a permutation group theorist.

By the mid-1970s, there was a feeling that the classification of finite simple groups might actually be completed, and maybe we would need to find something else to do.

John McDermott came to Oxford to give a seminar, and asked:

### Question

*Which infinite permutation groups  $G$  on  $\Omega$  have the property that they are **highly homogeneous** (transitive on the set of  $k$ -element subsets of  $\Omega$ ) for all  $k$  but not **highly transitive** (that is, they fail to be transitive on the set of ordered  $k$ -tuples of distinct points for some  $k$ )?*

## Three groups

I was able to answer the question:

### Theorem

*If  $G$  is highly homogeneous on the infinite set  $\Omega$ , then one of the following holds:*

- ▶ *there is a linear order on  $\Omega$  preserved or reversed by  $G$ ;*
- ▶ *there is a circular order on  $\Omega$  preserved or reversed by  $G$ ;*
- ▶  *$G$  is highly transitive.*

If  $\Omega$  is countable, then a linear order with highly homogeneous automorphism group must be dense and without endpoints, and hence isomorphic to  $(\mathbb{Q}, <)$ , by Cantor's theorem. Similarly a countable circular order admitting such a group must be isomorphic to the complex roots of unity.

So good things come in threes:

- ▶  $A$ , the group of order-preserving maps on  $\mathbb{Q}$ ;
- ▶  $C$ , preserving the circular order on the roots of unity;
- ▶  $S$ , the symmetric group.

This theorem is now regarded as the founding document in the study of reducts of countable homogeneous structures (in model theory), but I still feel there is more to it than that . . .

## Richard Thompson's groups

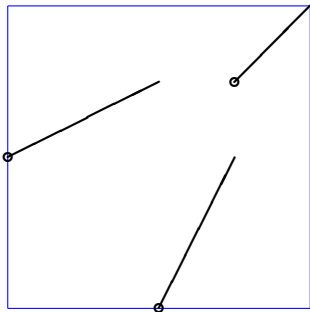
Earlier, Richard Thompson had constructed a remarkable triple of groups, which now go by the names  $F$ ,  $T$  and  $V$ . (I have no idea why.) They have many descriptions; here is one.

- ▶  $F$  is the group of piecewise-linear order-preserving permutations of the closed interval  $[0, 1]$  which are differentiable at all but finitely many dyadic rationals and has derivative a power of 2 on each interval.
- ▶  $T$  is the group of piecewise-linear order-preserving permutations of the circle  $S^1$  (regarded as the interval  $[0, 1]$  with endpoints identified) satisfying the same conditions as before.
- ▶  $V$  is the group of right-continuous bijections of  $S^1$  (as above) satisfying the same conditions as before.

Thus  $V$  is order-preserving on each interval, but there can be breaks in its graph.

## An example

Here is an element of  $V$ , taken from the exposition by Cannon, Floyd and Parry.



## Cantor space

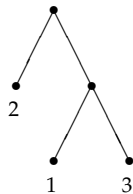
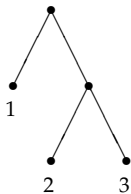
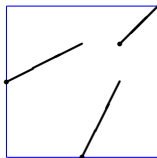
Such maps are homeomorphisms of the space remaining when we remove the dyadic rationals from the unit interval.

Points of this space have infinite dyadic expansions, and so can be identified with paths in the infinite binary tree. This is a realisation of the Cantor space.

Elements of  $V$  can be thought of as follows: take a finite subtree of the infinite binary tree; move the "branches" below the leaves of this finite subtree around arbitrarily, without any movement within these sets, to the corresponding branches in another finite subtree with the same number of leaves.

Here is the example seen earlier, realised in this form.

## Example revisited





## Properties

Thompson's groups have some remarkable properties. One good source of information is the set of notes by Cannon, Floyd and Parry.

For example,

- ▶  $F$  and  $T$  are  $FP_\infty$  groups,  $H^*(F, \mathbb{Z}F) = H^*(T, \mathbb{Z}T) = 0$ ;
- ▶  $T$  and  $V$  are finitely presented infinite simple groups;
- ▶  $F$  is not elementary amenable, but it is not known whether or not it is amenable.

I won't define all these things – they are here to show that these groups are important!

## Historical note

I first learned about Thompson's groups in lectures by Graham Higman. I do not know who invented what.

Assuming the Axiom of Choice, an infinite set  $X$  is bijective with its Cartesian square  $X \times X$ . The bijection can be defined by two unary functions  $\alpha$  and  $\beta$  on  $X$ , and its inverse by a binary function  $\lambda$ , satisfying the conditions (in postfix notation)

- ▶  $xy\lambda\alpha = x, xy\lambda\beta = y;$
- ▶  $z\alpha z\beta\lambda = z.$

The set of algebras with signature  $(0, 2, 1)$  satisfying these laws is a variety, so has free algebras.

$V$  is the automorphism group of the 1-generator free algebra in this variety.

Higman showed that, doing the same for a bijection between  $X$  and  $X^n$ , the automorphism group of the  $r$ -generator free algebra is a finitely presented infinite group  $G_{n,r}$ , which is simple if  $n$  is even and has a simple subgroup of index 2 if  $n$  is odd.

It is possible to match up both this definition and the earlier one with yet a third definition (much more practical and useful) involving transformations of a forest of  $r$   $n$ -ary trees (one binary tree in the case of the group  $V$ ).

Since the paths of such a tree are the points of Cantor space, this gives us an action of the group (by homeomorphisms) on the Cantor space (which can also be seen directly from the first definition, removing the dyadic rationals (the possible points of discontinuity) from the interval).

I think it is likely that, if Higman had mentioned the relationship with order-preserving permutations of lines and circles, I would have become more interested in these groups than I actually did at the time!

## Enter symbolic dynamics

Dynamicists have been interested in the shift map on sequences over a finite alphabet for a long time. The seminal paper on symbolic dynamics by Morse and Hedlund was published in the *American Journal of Mathematics* in 1938.

For example, the map  $x \mapsto 2x \pmod{1}$  induces the shift map on the space of binary sequences (identified with the base-2 “decimal expansion” of the argument  $x$ ): the leading digit drops off, and the others move up one.

This is also connected with Thompson’s groups!

## Shifts and automorphisms

This is based on the work of Bleak, Maissel and Navas. These authors show, among other things, that the outer automorphism group  $\text{Aut}(G_{n,r})$  of the Higman–Thompson group  $G_{n,r}$  is independent of  $r$ , and is a specific subgroup of the automorphism group of the one-sided shift over an alphabet of  $n$  symbols.

These authors also give a description of the elements of this group by means of transducers (see below).

I now turn to recent joint work with Collin Bleak in St Andrews, in which we also show that  $\text{Out}(G_{n,r})$  is the full automorphism group of the one-sided shift.

# Automata

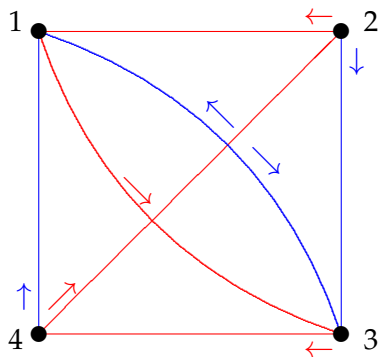
An **automaton** is a black box which can be in any one of a finite set of internal states. It can read a symbol from a fixed finite alphabet  $A$ , upon which it undergoes a state transition. This can happen repeatedly.

We can represent an automaton by a directed graph. The vertices of the graph are the states, and the edges are labelled with the elements of the alphabet. If the automaton is in state  $s$  and reads a symbol  $a$ , it moves to the state  $t$  for which an arc labelled  $a$  goes from  $s$  to  $t$ .

Our automata are **deterministic**: that is, there is exactly one edge with each possible label *leaving* each vertex.

That is it: there are no accept states, and no language is recognised by the automaton (unlike in most applications of automata theory in computer science).

## An example



You can check that (Blue, Red, Blue) takes you to room 1 no matter where you start.

We say that this automaton is **synchronizing**, and the above sequence is a **reset word** for it.

# Transducers

A **transducer** is an automaton which writes symbols as well as reading them.

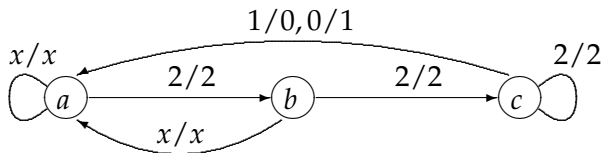
Each edge of a transducer carries two symbols  $a/b$ ; if it is in state  $s$  and reads symbol  $a$ , then it traverses this edge and writes symbol  $b$  to its output tape.

More generally, a transducer could write a finite string of symbols (possibly the empty string) at each step.

Thus, a transducer, in a given state and reading the first symbol of an infinite sequence, will write out a (potentially) infinite sequence, so inducing a map on the Cantor space of infinite sequences. The automorphisms of the shift referred to earlier are maps induced by transducers in this way.



## An example



Note:  $x \neq 2$

Applied to the infinite string  $222122202221 \dots$ , starting in state  $a$ , this transducer writes  $222022212220 \dots$

## Finitely determined automata

We say that an automaton is  *$k$ -determined* if every word of length  $k$  is a reset word for it. In other words, when it reads  $k$  symbols, the state it is in depends only on the symbols read, and not on the state it was in before reading them.

The automata involved in automorphisms of the Higman–Thompson groups turn out to be finitely determined. So we need to examine these further.

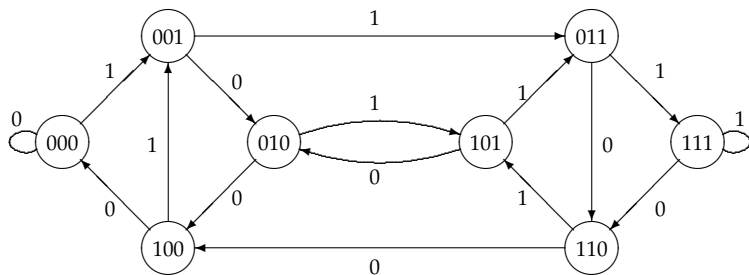
## De Bruijn graphs

The **de Bruijn graph**  $DB_{n,k}$  with word length  $k$  over an alphabet  $A$  of size  $n$  is defined as follows.

- ▶ The vertex set consists of all the words of length  $k$ .
- ▶ There is an edge from  $x_1x_2 \dots x_k$  to  $x_2x_3 \dots x_{k+1}$ , whose label is the  $(k+1)$ -tuple  $x_1x_2 \dots x_kx_{k+1}$ .

Since we want single symbols as labels, we will use just  $x_{k+1}$  for this edge.

Here is the de Bruijn graph  $DB_{2,3}$ .



These graphs were originally used to construct universal sequences. A **de Bruijn sequence** over  $A$  is a cyclic sequence of length  $n^{k+1}$  over  $A$ , with the property that each word of length  $k + 1$  occurs precisely once as a (consecutive) subsequence. The existence of de Bruijn sequences is immediate from the fact that the de Bruijn graph is strongly connected and has in-degree and out-degree equal, so is Eulerian; a Eulerian cycle gives the required sequence.

(An **Eulerian cycle** is a cycle passing once through each directed edge, in the correct direction; a directed graph has an Eulerian cycle if and only if it is strongly connected and has the in-degree of each vertex equal to its out-degree – an obvious necessary condition).

For us the crucial property (from which the strong connectedness follows) is that, regarded as an automaton over the alphabet  $A$ , the de Bruijn graph  $DB_{n,k}$  is  $k$ -determined.

## Foldings of de Bruijn graphs

A **folding** of  $DB_{n,k}$  is an equivalence relation  $\equiv$  on the vertex set with the property that, if  $v \equiv w$ , then for any symbol  $a$ , the vertices obtained by moving along edges labelled  $a$  from  $v$  and  $w$  are also equivalent.

- ▶ The quotient of  $DB_{n,k}$  by a folding is a  $k$ -determined automaton.
- ▶ Every  $k$ -determined automaton (in which every state is reachable) arises in this way.

So in order to study  $k$ -determined automata over  $A$ , we simply have to study foldings of  $DB_{n,k}$ .

## Counting foldings

*“I count a lot of things that there’s no need to count,” Cameron said. “Just because that’s the way I am. But I count all the things that need to be counted.”*

*Richard Brautigan, The Hawkline Monster: A Gothic Western*

If we really understand foldings, we should be able to count them. Let  $F(n, k)$  be the number of foldings of  $DB_{n,k}$ . Trivially,  $F(n, 1)$  is the Bell number  $B(n)$  (the number of partitions of the alphabet).

$n \setminus k$	1	2	3	4
2	2	5	30	1247
3	5	192	?	?
4	15	78721	?	?
5	52	519338423	?	?

## Word length 2

All but one of the results in the table were found by brute-force computation. However, we have found a formula for  $F(n, 2)$ :

### Theorem

Let

$$R(s, t) = \sum_{\pi} (-1)^{|\pi|-1} (|\pi| - 1)! \prod_{i=1}^{|\pi|} B(a_i s),$$

where  $\pi$  runs over all partitions of  $\{1, \dots, t\}$ ,  $|\pi|$  is the number of parts of  $\pi$ , and  $a_i$  is the size of the  $i$ th part. Then

$$F(n, 2) = \sum_{\pi} \prod_{i=1}^s R(m, a_i),$$

where  $\pi$  runs over all partitions of the  $n$ -letter alphabet,  $m$  is the number of parts of  $\pi$ , and  $a_i$  is the cardinality of the  $i$ th part for  $i = 1, \dots, m$ .

It looks complicated, but it allows the computation of  $F(20, 2)$ , a number of several hundred digits, in a second or so.

If you stare at the formula you will see the technique: Möbius inversion over the lattice of partitions of a set. But we haven't made it work for longer words yet!

**Möbius inversion** is a general technique for arbitrary partially ordered sets which generalises the Inclusion-Exclusion principle (the case for the lattice of subsets of a set). I have known the form of the Möbius function for the partition lattice for many years, but never before now did I have a chance to use it seriously.



## Counting automorphisms?

Even if we could count foldings of de Bruijn graphs, we are still some way off a good description of automorphisms. Among the things we would need to do are

- ▶ decide when two foldings give rise to isomorphic directed graphs, and count these;
- ▶ decide when two transducers have the same action, and so give rise to the same automorphism.

So we still have plenty more to do!