The Math Behind Futurama: The Prisoner of Benda

Reeve Garrett

May 7, 2013

The problem (informally)

Professor Farnsworth has created a “mind-switching” machine that switches two bodies, but the switching can’t be reversed using just those two bodies. Using this machine, some number of people got their bodies mixed up. Can we get all the minds back to their original bodies? If so, how?

An Introduction to Set Theory in 30 seconds

Definition:

A **set** is just a collection of objects we call **elements**. Whenever we mean "a is an element of the set A," we denote it by $a \in A$ as it makes the notation shorter.

A simple example:

We can list out the set of all people whose bodies were switched like this, calling the set A:
$A = \{Leela, Fry, Amy, Farnsworth, Bender, Washbucket, ...\}$.
But we can compactify this notation even more: if n people had their bodies switched, we can rename the people $1, ..., n$ and then denote this set by $A = \{1, 2, ..., n\}$.

Functions and Permutations

Definitions:

Given 2 sets X and Y, a **function** f from X to Y, which we often denote using the notation $f : X \to Y$, is a rule that assigns each element x of X to an element of Y we'll call $f(x)$. A function $f : X \to Y$ is called a **bijection** if the following two criteria hold: (1) For each $y \in Y$ there exists an $x \in X$ such that $y = f(x)$ and (2) if for two elements x_1 and x_2 of X we have $f(x_1) = f(x_2)$, then x_1 and x_2 are the same element. A bijection from a set X to itself is called a **permutation** on X.

Examples and Some Motivation

Example:

People and chairs

Examples:

The function sending an integer to itself plus one is a permutation on the integers. The function sending an integer to itself squared is not, as $(-3)^2 = 3^2$, so (2) fails.

Main point:

The scrambling of minds and bodies in the episode can be thought of as a permutation/bijection on the set of people whose bodies were switched by Farnsworth's machine.

Binary operations

Definition:

A **binary operation** on a set X is a rule $*$ that assigns each ordered pair (x, y) of elements of X to a new element of X we call $x * y$.

Examples:

$+$, $-$, $\times$ on the integers; function composition $\circ$ on the set of permutations $f : X \to X$.

Function composition

Take a permutation $f : X \to X$ and another permutation $g : X \to X$. We define a new function $g \circ f$ in the following way: To each $x \in X$, x is assigned to an element $f(x)$ by the function f, and the element $f(x)$ is assigned to an element $g(f(x))$ by the function g, and we call this element $(g \circ f)(x)$, which is in X. So, first we apply f to x, and then we apply g to that result, and this defines the element that x gets sent to by our "new" function $g \circ f$.

Mnemonic: Function composition is done *right to left*.

Example of function composition

Let $\mathbb{R}$ denote the set of real numbers $(2, \frac{1}{3}, \pi$, etc.). Consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ and $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x + 1$. We have $(f \circ g)(x) = (x + 1)^3$ and $(g \circ f)(x) = x^3 + 1$ for each real number x.

Consequence: In general, given two permutations f and g on a set X, $f \circ g \neq g \circ f$.

Groups

Definition:

A **group** is a set G with binary operation $*$ such that

(i) $*$ is an associative operation; i.e., $a * (b * c) = (a * b) * c$

(ii) there is an identity element with respect to $*$; that is, there exists some $e \in G$ such that for all $g \in G$, $e * g = g * e = g$

(iii) each element in G admits an inverse with respect to $*$; that, is for all $g \in G$, there exists a g^{-1} such that $g * g^{-1} = g^{-1} * g = e$

Examples:

The integers $\mathbb{Z}$ with respect to $+$, the nonzero fractions $\mathbb{Q}$ with respect to $\times$

The Permutation Group

Fact: The set of permutations on a set X, denoted $S(X)$, with binary operation $\circ$ (function composition) forms a group!

Verification:

(i) For all $f, g, h \in S(X)$ and for all $x \in X$, $((f \circ g) \circ h)(x) = (f \circ g)(h(x)) = f(g(h(x)))$ and $(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x)))$. So, $f \circ (g \circ h) = (f \circ g) \circ h$ for all f, g, h, and $\circ$ is associative.

(ii) We have the identity permutation id_X, and indeed, $id_X \circ f = f \circ id_X = f$ for all $f \in S(X)$.

(iii) Permutations are bijections and thus invertible. Given a permutation f, for any $y \in X$, there is a unique $x \in X$ such that $f(x) = y$. Define a new permutation f^{-1} by setting $f^{-1}(y) = x$. Thus, $f \circ f^{-1} = f^{-1} \circ f = id_X$.

Special notation for permutation groups on finite sets

When X is a finite set with n elements, we rename the elements $\{1, 2, ..., n\}$.

Each permutation f on X can be identified with a box of numbers like the following:

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix}$$

One interpretation of this is that it can be thought of as telling us "thing in person 1's body gets sent to person $f(1)$'s body," and so on.

With this new notation, we compose permutations like this, right to left:

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ f(1) & f(2) & \cdots & f(n) \end{pmatrix} \begin{pmatrix} 1 & 2 & \cdots & n \\ g(1) & g(2) & \cdots & g(n) \end{pmatrix} = \begin{pmatrix} 1 & 2 & \cdots & n \\ f \circ g(1) & f \circ g(2) & \cdots & f \circ g(n) \end{pmatrix}$$

Examples of composition:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

while

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

Moral:

Order of composition almost always matters!

Example and Notation for Transpositions

Example:

If I have a permutation f on the set $\{1, 2, 3\}$ that sends 1 to 2, 2 to 1, and 3 to itself (so, $f(1) = 2$, $f(2) = 1$, and $f(3) = 3$), I write the permutation as

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

A permutation like this that just interchanges two elements and keeps everything else in place is called a **transposition**.

Special notation for transpositions:

More generally, if in $\{1, ..., n\}$ we just switch the places of i and j, we denote the transposition simply as $(i\ j)$. So, in the above example, we write it as $(1\ 2)$.

Cycles

Transpositions are also called "2-cycles" because they cycle through two elements. We can extend this notation to arbitrary lengths. So, for instance, on the set $\{1, 2, 3, 4, 5, 6\}$, an example of a cycle would be (2 5 3 1), which tells us 2 goes to 5, 5 goes to 3, 3 goes to 1, and 1 goes to 2 (and everything else doesn't move). Using our box notation, this is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 1 & 4 & 3 & 6 \end{pmatrix}$$

Two cycles are called **disjoint** if they don't have any numbers in common, so (1 4 5) and (2 3 6 7) are disjoint, but (1 2 3) and (5 2 4 6) are not disjoint because 2 appears in both. It's easy to see that the order of composition doesn't matter for disjoint cycles because one cycle leaves everything involved in the other cycle unchanged.

An Important Theorem

Theorem

Any permutation on a finite set can be written as a composition of disjoint cycles.

Example that gives the main idea of the proof

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 5 & 1 & 8 & 3 & 7 & 4 & 6 \end{pmatrix}$$

can be decomposed into $(1\ 2\ 5\ 3)$ and $(4\ 8\ 6\ 7)$.

Consequence:

We only have to think of the problem posed in the episode in the case of cycles! Hurray!

Keeler's Theorem

Theorem:

Given the set $\underline{n} = \{1, 2, ..., n\}$ (the $\underline{\text{n}}$ is just my own abbreviation) and an arbitrary permutation on $\underline{\text{n}}$, if two additional elements x and y are "adjoined" to $\underline{\text{n}}$ (meaning we add x and y to make a new set $\underline{\text{n}} \cup \{x, y\}$), can be reduced to the identity permutation by applying a sequence of distinct transpositions of $\underline{\text{n}} \cup \{x, y\}$, each of which includes at least one of x or y.

Keeler's Theorem (ctd.)

Sketch of proof

Consider each of the disjoint cycles involved in the permutation one at a time. As a result, if a cycle has length k, we can temporarily rename it as $(1\ 2\ \cdots k)$. Then, fixing i between 1 and k arbitrarily, apply
$\mu = (x\ 1)(x\ 2)\cdots(x\ i)(y\ i+1)(y\ i+2)\cdots(y\ k)(x\ i+1)(y\ 1)$
and we get

$$\begin{pmatrix} 1 & 2 & \cdots & k & x & y \\ 1 & 2 & \cdots & k & y & x \end{pmatrix}$$

Repeat with other cycles and transpose x and y at the end if necessary.

Questions?

Thanks for showing up!

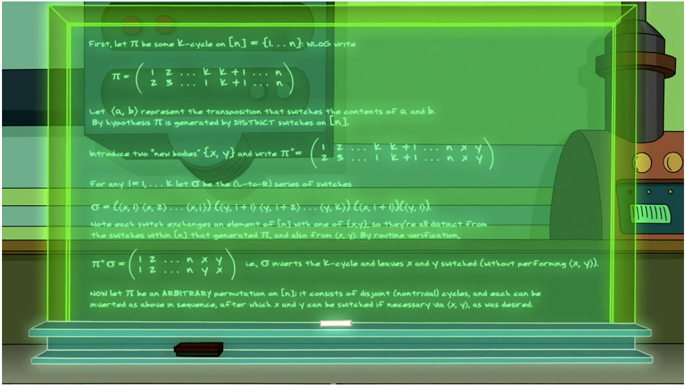

Figure: Image from the episode