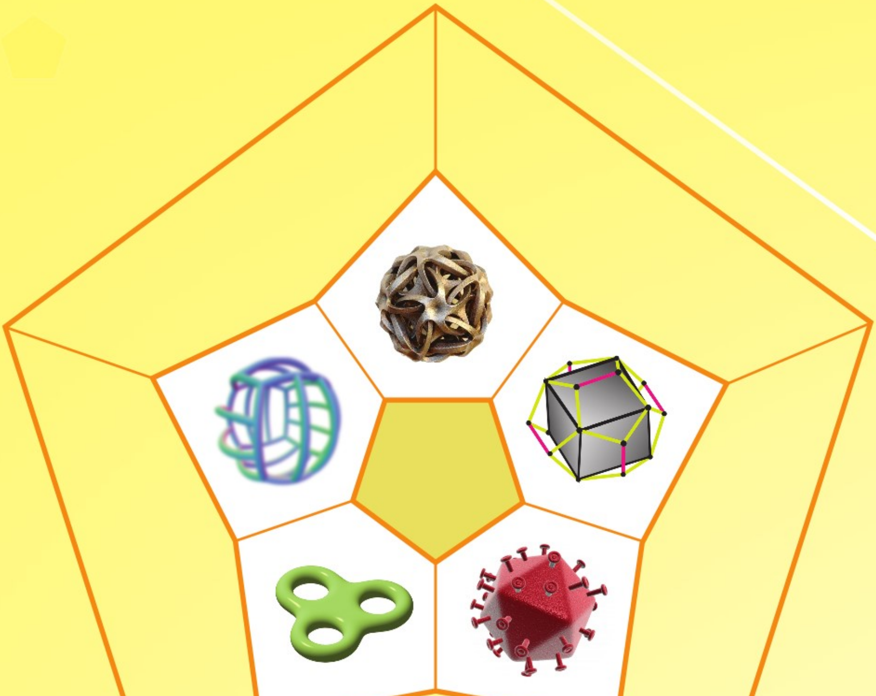
Kristopher Tapp

# Symmetry

## A Mathematical Exploration



Springer

Symmetry

Kristopher Tapp

# Symmetry

## A Mathematical Exploration

Springer

Kristopher Tapp
Department of Mathematics
Saint Joseph's University
5600 City Ave.
Philadelphia, PA 19131
USA
ktapp@sju.edu

ISBN 978-1-4614-0298-5 e-ISBN 978-1-4614-0299-2
DOI 10.1007/ 978-1-4614-0299-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011940690

Mathematics Subject Classification (2010): 00-01, 00A06, 00A09, 00A66, 00A99, 20-01, 51-01, 97G40, 97U20

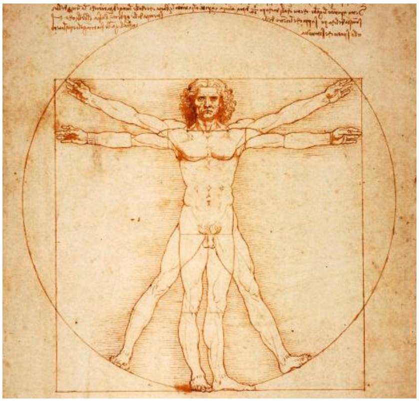


Printed on acid-free paper

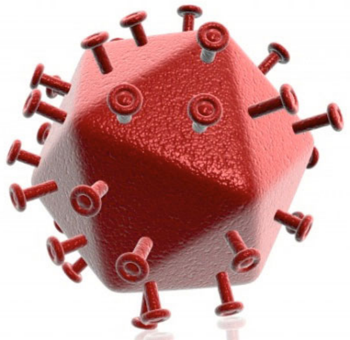
Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

Symmetry is a ubiquitous concept in mathematics and science. Certain shapes and images seem more symmetric than others, yet it is not immediately obvious how to best measure and understand an object's symmetry. In fact, the quest to more precisely understand symmetry has been a driving force in science and mathematics, and will form the central theme of this book. You will learn the ways in which mathematicians study the topic of symmetry.



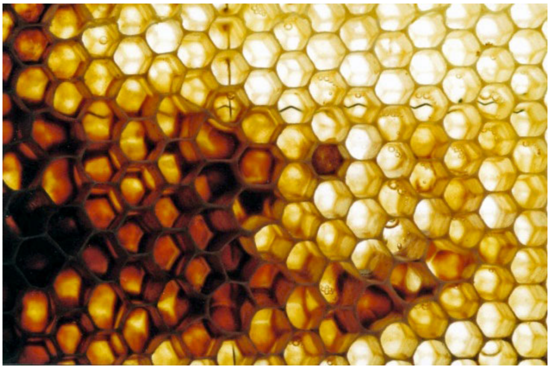
*Vitruvian Man* by Leonardo Da Vinci



An icosahedral HIV virus

If you are curious about the mathematical patterns underlying the symmetry that you observe in the physical world, then this book is for you. Why are honeycombs hexagonal? Why are bubbles spherical? Why did the HIV virus evolve its icosahedral shape? What is the shape of the universe, and how might this shape be related to the shape of a virus? How can one understand the symmetry of molecules or crystal formations? How might the symmetry in a painting enhance its artistic appeal? Parts of these answers are found in other disciplines – biology, chemistry, physics, and art – but the common thread is mathematics. Mathematics provides the tools to understand and classify the possible types of symmetry that objects may possess, which is a crucial prerequisite for addressing questions like those above.

No background beyond high school level algebra is required to read this book. The mathematical topics are drawn from diverse fields including graph theory, abstract algebra, linear algebra and topology, all of which are essential to rigorously study symmetry. Although some of these topics are advanced, the presentation in this book is intended to be precise and rigorous, yet accessible to a general audience. The only real prerequisite is that you discard any preconceived notions of what math is and is not, and begin this mathematical journey with an open mind and a willingness to begin actively doing what mathematicians do: discovering patterns, inventing precise language for discussing the mathematical principles underlying those patterns, forming conjectures, and eventually proving beautiful theorems.



Honeycomb photo by Ken Tapp

## Intended Audience

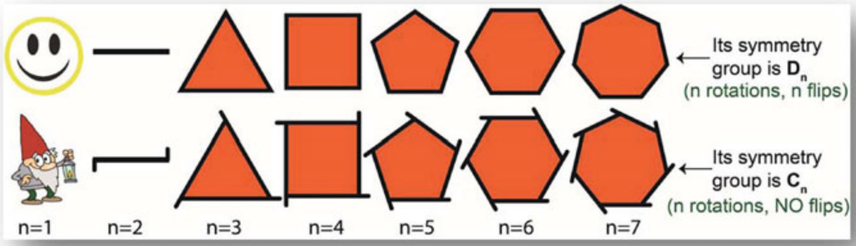
This book is primarily intended as a textbook for a one-semester math course for math or non-math majors, including humanities majors, with the goal of encouraging effective analytical thinking and exposing students to elegant mathematical ideas. It includes some of the topics which are commonly found in sampler textbooks, such as Platonic solids, Euler's formula, irrational numbers, countable sets, permutations and a proof of the Pythagorean Theorem. All of these topics serve a single compelling goal: to understand the mathematical patterns underlying the symmetry that we observe in the physical world. I hope that students from all majors will enjoy the many beautiful mathematical topics herein, and will come to better appreciate the powerful cumulative nature of mathematics as these topics are woven together into a single story about symmetry.

Instructor resources, including PowerPoint lectures and access to all images in the book, can be found at http://www.sju.edu/~ktapp/Symmetry

## Acknowledgments

I am delighted to thank Paul Klingsberg and Susan Ramee for suggesting many improvements to the exposition. I am also grateful to the following artists who generously permitted me to include their images in my book: Vladimir Bulatov, Robert Fathauer, Brian Sanderson, Paul Söderholm and Ken Tapp.

# Table of Notation




$\mathbf{D}_n$ = the *n*th dihedral group = the symmetry group of a regular *n*-gon. It has 2*n* total members (*n* rotations and *n* flips).

$\mathbf{C}_n$ = the *n*th cyclic group = the symmetry group of an *oriented* regular *n*-gon. It has *n* total member (all rotations). A shorthand notation for these n rotations is {0, 1, 2, …, *n*–1}.

$\mathbf{P}_n$ = the *n*th permutation group = the collection of all permutations of *n* ordered things. It has *n*! total members.

$\mathbf{A}_n$ = the *n*th alternating group = the collection of all *even* permutations of *n* ordered things. It has half the size of $\mathbf{P}_n$.

**Z** = {… –3, –2, –1, 0, 1, 2, 3, …} = the set of all integers.

**Q** = the set of all rational numbers (fractions like 8/5 and –3/7).

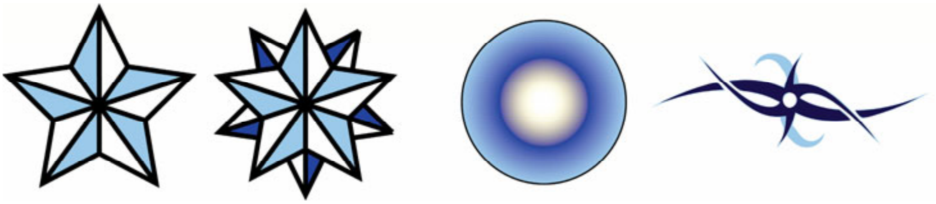
**R** = the set of all real numbers (rational and irrational numbers).

$\mathbf{R}^n$ = *n*-dimensional Euclidean space. For example, $\mathbf{R}^2$ is called *the plane* and $\mathbf{R}^3$ is called *space*.

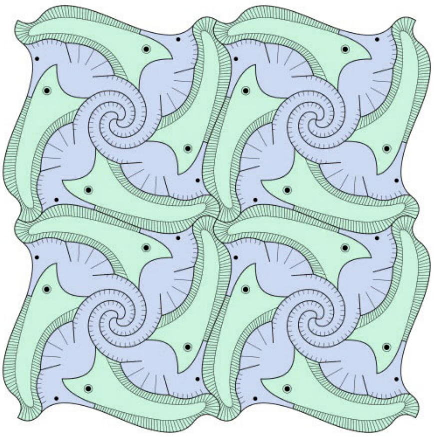
# Table of Contents

# 1. Introduction to Symmetry

Our journey starts with the question: what does “symmetry” mean? Look at the following four objects, and rank them from the most symmetric to the least symmetric:



How do you interpret this question in a manner which is precise enough to lead you to a justifiable ranking of the four objects? And how symmetric is the following painting by Robert Fathauer?



*Seahorses and Eels* by Robert Fathauer
*http://members.cox.net/fathauerart/*

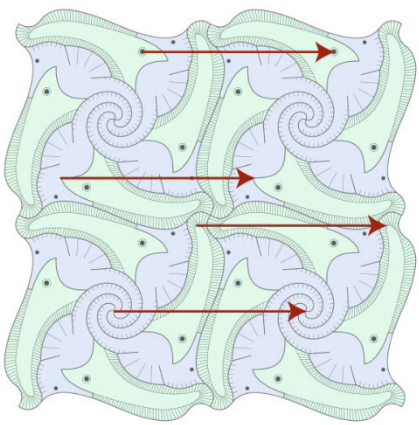
To answer any of these questions, we must first make the questions more precise.

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_1,

## A Precise Definition of "Symmetry"

Most people would agree that *Seahorses and Eels* looks symmetric, perhaps because it contains repeated images. But symmetry involves more than just repeated images. A haphazard arrangement of green eels and purple seahorses would not look nearly as symmetric. Fathauer arranged his seahorses and eels together like jigsaw puzzle pieces, so that the pattern of neighbors surrounding one eel is the same as the pattern surrounding any other.

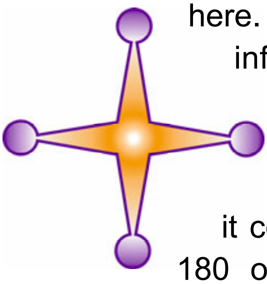
To phrase this idea more precisely, let us imagine that the pattern is painted onto an infinite glass wall that extends indefinitely up, down, right and left. Imagine that the painted pattern extends indefinitely so as to cover the whole infinite wall, which requires infinitely many seahorses and infinitely many eels. This infinite painting looks exactly the same from many different positions. If the viewer is positioned in front of the eye of one right-facing eel, then what she sees is exactly the same as if she were positioned in front of the eye of another right-facing eel.



A translation symmetry

Here is an equivalent way to say the same thing, rephrased in terms of moving the glass wall rather than moving the viewer: there are many ways in which the glass wall could be moved/ repositioned so that the painted image looks exactly the same before and after the re-positioning. For example the wall, together with the pattern painted onto it, could be translated (which means slid) so that each right-facing eel moves one position to the right. This translation is called a symmetry of the infinite painting because a viewer who closes her eyes while the wall is moved, could not, after she opens her eyes, detect that any change has occurred. This translation is encoded by the length (about an inch) and direction (right) of the red arrow pictured above. Several copies of the red arrow are included to demonstrate that each composition element (the eye of a right-facing eel, the tail of a down-facing eel, the center of the purple tail spiral, etc.) moves exactly onto an identical element. That is why a viewer would not detect the change.

This way of thinking about symmetry applies equally well to other objects, including the orange and the purple star pictured here. Imagine this star image is painted onto our infinite glass wall. Again, there are several ways in which the wall could be repositioned/moved so that the image looks exactly the same before and after the repositioning. For example, it could be rotated 90° about the star's center, or 180 or 270°. Each of these motions is called a symmetry of the star image.



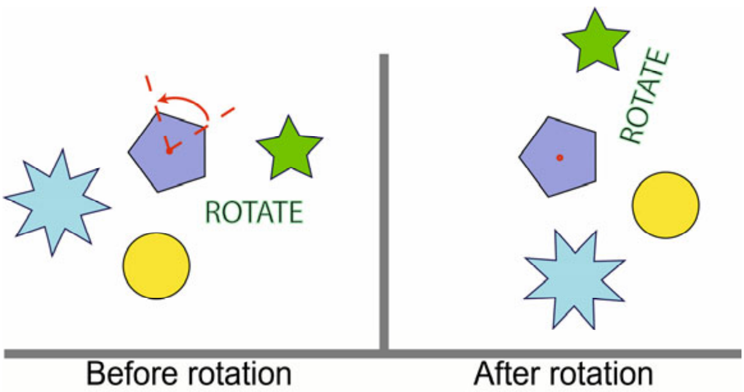
Here is our first attempt at formulating a precise definition of the word "symmetry":

> **DEFINITION:** A <u>symmetry</u> of an object in the plane is a rigid motion of the plane that leaves the object apparently unchanged.

In the above discussions, the "plane" was represented as an infinite glass wall, the "object" was represented as a painting on the wall (of a star or an infinite pattern), and a "rigid motion" meant a moving or repositioning of the glass wall, like a rotation or a translation. What does it mean for a rigid motion to "leave the object apparently unchanged"? It means that, if a viewer was to close her eyes during the repositioning, she would not detect a difference; the object would look exactly the same when she opened her eyes.

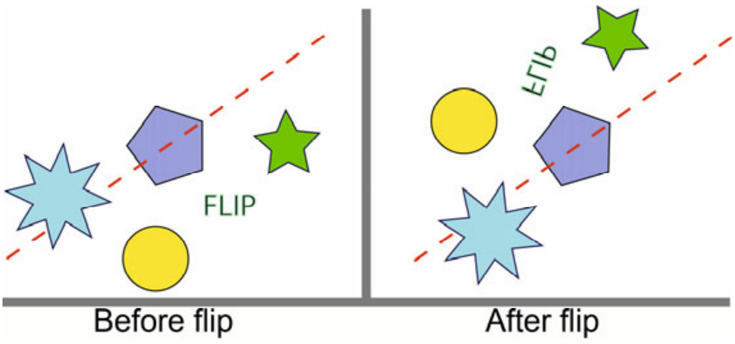
Precise language is crucially important in mathematics! Our above definition of symmetry cuts straight to the underlying reason that objects intuitively appear symmetric (they look the same from many positions and/or contain repeated images), but in a manner which is precise enough to form a foundation for a rigorous mathematical investigation of symmetry. To really pull this off, we will eventually require a more precise definition of the term “rigid motion”. But for now, it will be enough to think of a rigid motion as a moving/repositioning of the glass wall, like a rotation or a translation. A rigid motion may NOT break, bend, stretch, compress or otherwise distort distances on the glass wall (you cannot use a glass cutter or a blow torch).

A rigid motion is always a motion of the whole plane (the whole glass wall); one may then ask whether it is a symmetry of any object in the plane. For example, in the illustration below, the 72° rotation about the red point is a rigid motion of the plane that is a symmetry of the purple pentagon, but is not a symmetry of the other shapes.



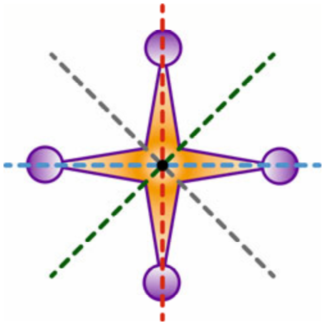
We emphasize that a rigid motion is completely determined by its effect on each point of the plane; that is, two motions that do the same thing to every point are considered the same motion. For example, a rotation by 90° is the same as a rotation by 90+360 = 450° about the same point. When enumerating the symmetries of an object, we would NOT list both a 90 and a 450° rotation, or any other such redundancies. What matters is the effect of a motion, not the motion itself.

One type of rigid motion that we have not yet considered is a flip over a line. Visualize a flip over a line as achieved by flipping the glass wall over to expose its back surface (its underside). Points along the line remain in their original position, while points on one side of the line flip over to the opposite side of the line. Imagine that the plane is a completely transparent glass wall, so that any image in the plane shows through the back surface of the glass, and looks reversed after the flip is performed. For example, in the illustration below, the flip over the red line is a rigid motion of the plane that is a symmetry of the purple pentagon but is not a symmetry of the other shapes.




The words <u>flip</u> and <u>reflection</u> are synonymous; we will henceforth use these terms interchangeably.

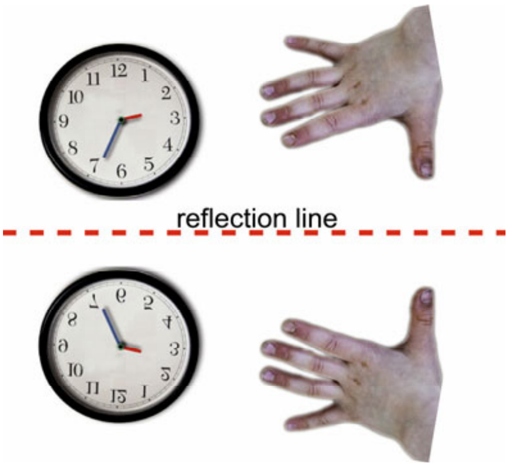
To better understand why the previously illustrated star image appears symmetric, we will list all of its symmetries. There are three obvious symmetries; namely rotations about its center point by 90, 180, and 270° *(rotation angles are always counterclockwise in this book)*. A fourth valid symmetry is called <u>the identity</u>. This is the motion that does nothing; it leaves every point of the plane in its original position. The identity can be considered a rotation (by 0°) or a translation (by zero distance). The identity is the only rigid motion that is a symmetry of every object. Thus, the star has four rotation symmetries. The star can also be flipped over any of the four colored lines illustrated here. In summary, the star has four rotation symmetries and four reflection symmetries, giving exactly eight symmetries in total. The number of symmetries which an object has provides a measurement of how symmetric that object is.


The star's 4 reflection lines

## Types of Symmetries and Types of Objects

To describe to me your favorite rotation or translation or flip, what information must you give me? A rotation is specified by its center point and its (counterclockwise) angle. A translation is specified by the length and direction of a single arrow. A flip is specified by its reflection line.

Rotations and translations are called "proper" rigid motions. Flips are called "improper" rigid motions. The intuitive difference is that improper motions leave the plane's underside facing the viewer. A more precise explanation of this difference is obtained by comparing how proper and improper motions affect a right hand or a clock. In the illustration on the right, flipping the top image over the red line transforms the right hand into a left hand and the clock into a “counterclock” (a clock that turns counterclockwise).




A flip reverses orientation

**DEFINITION:** A rigid motion is called proper if it preserves orientation, which means that after the motion is applied, an image of a right hand still looks like a right hand and a clock still looks like a clock. It is called improper if it reverses orientation, which means that it turns a right hand into a left hand and a clock into a counterclock.

An even more precise definition of proper/improper will be discussed later when we learn about matrices. For now, let us turn our attention to another intuitive concept which needs to be described more precisely. We previously imagined that the pattern in the Seahorses and Eels painting was extended infinitely up, down, right and left, so the resulting object is “unbounded”. On the other hand, the star image did not extend infinitely in any direction; we could fit the entire star image into a frame, so it is called “bounded”. We make this distinction precise by focusing, not on the imprecise “extended infinitely” verbiage, but instead on more precise issue of whether the object can be framed (say by a square frame):

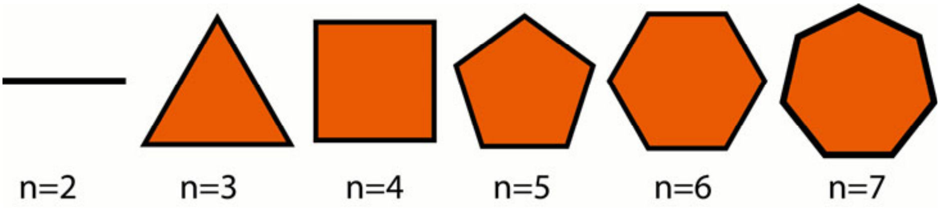
**DEFINITION:** An object in the plane is called bounded if it is fully contained in some square in the plane. Otherwise it is called unbounded.

The meaning would remain unaltered if the word "square" were replaced by "circle" or "pentagon" or many other possibilities. If an image can be framed by one of these frame shapes, then it can be framed by all of them.

In the study of symmetry, the most important bounded objects in the plane are the "regular polygons."

**DEFINITION:** The regular *n*-sided polygon (also called the regular *n*-gon) is the shape in the plane enclosed by *n* equal length straight sides, assembled so that all *n* of its angles are equal.

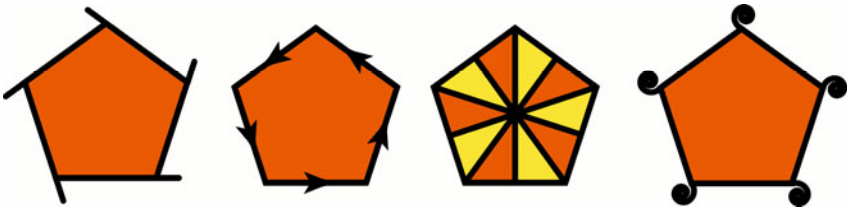
Thus, a regular 3-gon means an equilateral triangle, a regular 4-gon means a square, a regular 5-gon means a pentagon, a regular 6-gon means a hexagon, and so on.




Regular polygons

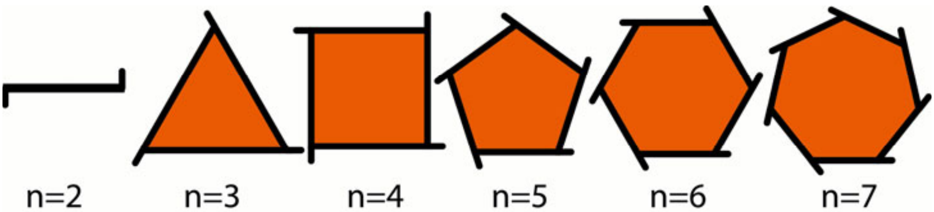
The two sides of the 2-gon lie on top of each other (because they meet at angles of 0°), so the 2-gon looks like a line segment.

Notice that each of these regular polygons has both rotation and reflection symmetries. Can you think of a way to orient each of these polygons, which means to alter it in such a way that its rotation symmetries are preserved but it no longer has any reflection symmetries? For example, here are a few artistic ways to orient the pentagon.



Oriented pentagons

There are many other possibilities; whichever you choose, the result is called an oriented pentagon. Each oriented pentagon pictured above has five rotation symmetries but NO reflection symmetries. Do you see why? A reflection would reverse the issue of whether it appears to spin clockwise or counterclockwise, and would therefore not be a symmetry. Since the first method is the simplest, we will use it to orient the other regular polygons:




Oriented regular polygons

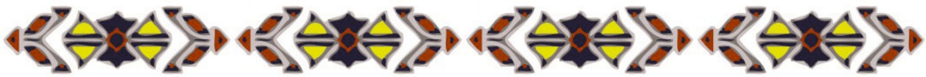
Here is the general definition:

**DEFINITION:** An object in the plane is called oriented if it has NO improper symmetries.

You can always detect that an oriented object has been flipped. A clock is oriented because flipping it would make it look like a counterclock. Similarly, each oriented regular polygon above appears to spin counterclockwise, but would appear to spin clockwise after being flipped. The knotted blue object pictured here is oriented; hold it up to a mirror, and notice how its over/under crossing pattern differs from that of its mirror image. A flip would make it look like its mirror image, so a flip could not be a symmetry.

An oriented object

There are two types of unbounded objects that are classically important within the study of symmetry. First, a wallpaper pattern intuitively means an unbounded pattern that extends infinitely in all directions (left, right, up, and down) according to some organized scheme. The Seahorse and Eels painting is a wallpaper pattern (after being indefinitely extended). Second, a border pattern (also called a Frieze pattern) means an unbounded pattern that only extends infinitely along one line (usually the *x*-axis). For example, if the following pattern is extended infinitely to the right and left, then the result is a border pattern:



A border pattern

Border patterns are usually drawn horizontally as above, so they extend infinitely to the right and left, but not up or down. When positioned like this, all of the pattern’s translation symmetries are in directions parallel to a horizontal line. This observation helps us to formulate a more precise definition.

**DEFINITION:** An unbounded pattern in the plane that has at least one translation symmetry (besides the identity) is called

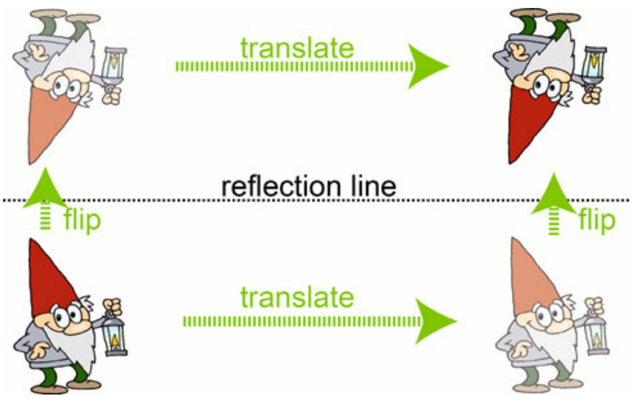
a border pattern if all its translations are parallel to a single line.

a wallpaper pattern otherwise.

The previously illustrated border pattern has many types of symmetries. You can translate it any number of positions to the right or left. You can reflect it over the horizontal center line. There are also vertical lines over which you can reflect it. If you perform any pair of the above-mentioned symmetries, one after the other, the result will also be a symmetry. For example, if you translate it any number of positions to the right or left and then reflect it over the horizontal center line, then the result is aptly called a glide reflection.

**DEFINITION:** A glide reflection means the result of performing a translation (other than the identity) followed by a reflection over a line that is parallel to the direction of the translation.

It does not matter which you do first: translate or reflect. In the illustration below, either order has the same effect of moving the bottom-left gnome to the top-right position.




A glide reflection translates and flips
Gnome image (used here and elsewhere) created by Paul Söderholm, www.gnurf.net.

Can you invent a border pattern with a glide reflection symmetry that has the peculiar property that the reflection and translation out of which it is built are not themselves symmetries of the border pattern? An answer is found in an exercise at the end of this chapter.

## The Classification of Plane Rigid Motions

We began with vague intuitive notions of the word "symmetry." A symmetric object often contains repeated images, and often looks the same from many positions. Based on these, we formulated a mathematically precise definition: a symmetry of an object is a rigid motion of the plane that leaves the object apparently unchanged. We then formulated precise definitions of other terms: "bounded", "proper", "oriented", "border pattern", and "wallpaper pattern". This provides us with a vocabulary for more precisely discussing symmetry. In the remainder of the book, this precision will serve us well. It will allow us to ask and answer many precise questions, and eventually to prove beautiful theorems about the possible types of symmetries that objects may have. In this book, definitions are placed in green boxes and theorems are placed in blue boxes.

What is still missing? Well, we defined a "symmetry" using the term "rigid motion" but we have not yet precisely defined the term "rigid motion". Rather, we have relied on an intuitive feeling for this concept. When it becomes necessary, we will eventually give a more precise definition of "rigid motion." To help us get by for now, we mention the following (which our eventual precise definition will allow us to prove):

**CLASSIFICATION OF PLANE RIGID MOTIONS (VERSION 1):**

Every proper rigid motion of the plane is a translation, a rotation, or a rotation followed by a translation.

Every improper rigid motion of the plane is a flip or a flip followed by a translation.

In other words, there are no rigid motions other than the types that we have already considered (and combinations thereof). You may take this classification as your definition of rigid motion for now, if you like.

The story is even simpler for rigid motions that are symmetries of a *bounded* object. The symmetries of a bounded object include only flips and rotations (no translations). In fact:

**THE CENTER POINT THEOREM:** Any bounded object in the plane has a "center point" such that:

**(1)** Every proper symmetry is a rotation about this center point.

**(2)** Every improper symmetry is a flip over a line through this center point.

You might think of an object's center point as a balancing point; if you cut the object out of cardboard and wish to balance it on your finger tip, this is the correct place to position your finger.

## Exercises

*Challenge problems are designated* .

(1) Which are proper and which are improper:

1. A proper symmetry followed by an proper symmetry
2. A proper symmetry followed by an improper symmetry
3. An improper symmetry followed by an improper symmetry

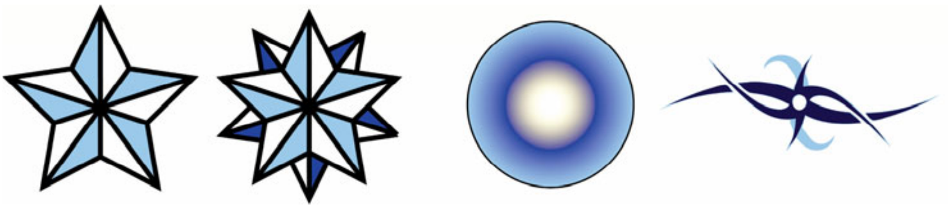
(2) How many symmetries does each capital letter in the English language have:

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z.

(3) How many symmetries does the $n$-gon have for each of the values $n = 2, 3, 4, 5, 6, 7$ (these polygons are illustrated in the chapter)? Guess a general formula for the number of symmetries of an $n$-gon. What about an oriented $n$-gon?

(4) Draw a wallpaper pattern with NO improper symmetries.

(5) How many symmetries does each object have? Which object has the most? The least?



(6) Any capital letter in the English language can be used to create a border pattern like this:

…A A A A A A A A….
…B B B B B B B B …
…C C C C C C C C …

For each of the 26 letters, decide whether the resulting border pattern (a) contains reflections across any horizontal lines, (b) contains reflections across any vertical lines, and (c) contains rotations (other than by 0°).

(7) Consider the following border pattern to be extended infinitely to the right and left.

Characterize all of its translation, rotation, reflection, and glide reflection symmetries. What if the pattern was built from As rather than Rs?

(8) Draw two different objects that have exactly the same collection of symmetries.

(9) How many bounded objects can you think of that have infinitely many symmetries?

(10) What do you think is the most symmetric object in the plane?

(11) Make sketches of several bounded objects that have interesting collections of symmetries. Try to sketch a bounded object whose collection of symmetries is significantly different from that of an oriented or non-oriented polygon.

(12) "If an object has any translation symmetries (other than the identity), then it must have infinitely many translation symmetries." Explain why this statement is true.

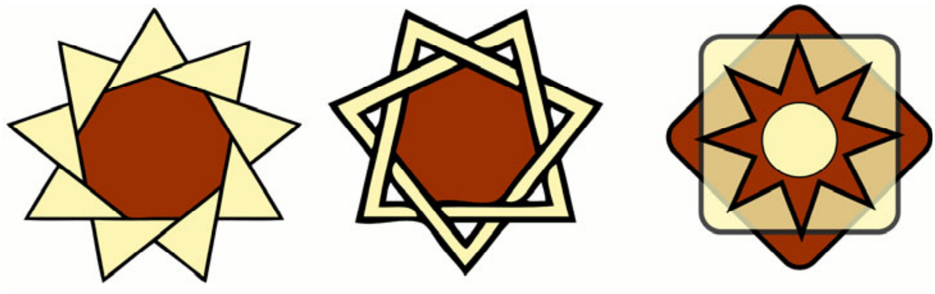
(13) "If an object has any translation symmetries (other than the identity), then it must be an unbounded object." Prove this statement from scratch (without using The Center Point Theorem). *HINT: Visualize the object as painted on the (glass)*

*plane. Where will a single drop of paint be moved as one translation symmetry is repeatedly performed?*

(14) “A bounded object could never have any translation symmetries other than the identity.” Explain how this statement is related to the statement in the previous problem?

(15) Do you think that a circle can be oriented? In other words, do you think there is an oriented object that has the same proper symmetries as the circle? Guessing is fine – you do not need to prove your answer.

(16) Count the proper and improper symmetries of each star pictured below. Which are oriented?



(17) Identify some rotation symmetries of the previously pictured *Seahorses and Eels* wallpaper pattern.

(18) Count the symmetries of the *Seahorses and Eels* painting considered at face value as a bounded image (not extended indefinitely into a wallpaper pattern).

(19) Perform a web image search for “M.C. Escher symmetry”. Many of Escher’s paintings are wallpaper patterns (if you imagine them infinitely extended). How many different angles of rotational symmetry can you find among his wallpaper pattern paintings? Identify some patterns that are oriented and some that are not.

# 2. The Algebra of Symmetry

You learned as a child how to count, and thus became familiar with the numbers, but that was not nearly enough. To really understand numbers, you also needed to learn how to perform algebraic operations on numbers, such as addition and multiplication. Addition and multiplication are algebraic ways of combining two numbers to get a number back as the answer. These algebraic operations greatly enriched your ability to understand, appreciate, and work effectively with numbers.

It is the same with symmetries. In the previous chapter, you listed all symmetries of an object (such as a triangle or a square). But just listing them is not nearly enough. To really understand, appreciate, and work effectively with the collection of symmetries that you listed, you need to learn a crucial algebraic operation called "composition." It allows you to combine two symmetries of your object to get back a symmetry of that object as the answer.

**DEFINITION:** If A and B are rigid motions of the plane, then A*B denotes the rigid motion obtained by first performing B and then performing A. It is called the <u>composition</u> of A with B.

If A and B are both symmetries of an object (say a triangle or rectangle or wallpaper pattern), then A*B is also a symmetry of that object. Thus, composition is an algebraic operation on the collection of symmetries of that object. It combines two symmetries of the object to get back a symmetry of the object as

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_2, 

the answer. Think of "A*B" as meaning "A following B" or "A performed after B." It is important to keep the order straight – it is the opposite of what you might have expected.

## Cayley Tables

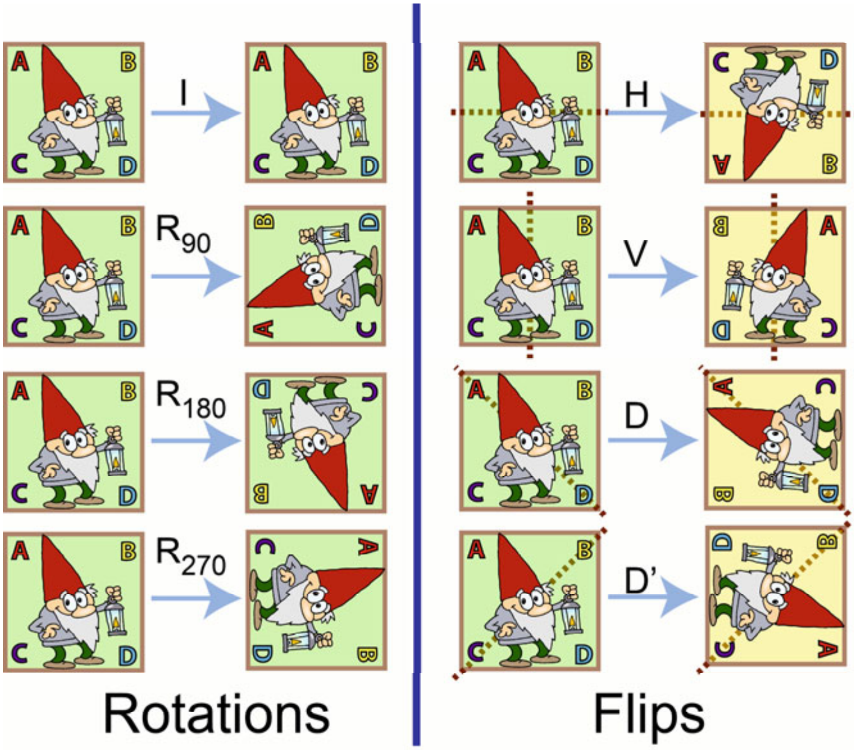
As a child, you became familiar with the algebraic operations of addition and multiplication by memorizing tables. Similarly, you will now study the algebraic operation of composition by building a table that exhibits the result of composing any pair of symmetries.

Let us start with a square. Its eight symmetries are:

$$\{\mathbf{I}, \mathbf{R}_{90}, \mathbf{R}_{180}, \mathbf{R}_{270}, \mathbf{H}, \mathbf{V}, \mathbf{D}, \mathbf{D'}\}$$

where **H** means horizontal flip, **V** means vertical flip, **D** and **D′** mean the two diagonal flips, and **R** means a counterclockwise rotation by the subscripted angle. The illustrations below show the effect of these eight symmetries on a square whose corners are labeled A, B, C, D and whose center is decorated with a picture of a gnome. The front of the square is green and the back is yellow.

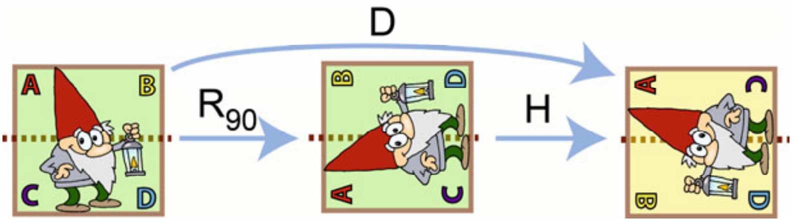
These illustrations are followed by a table which exhibits the composition of any pair of these symmetries. This table is called a Cayley table for the square (or a Cayley table for the symmetries of the square). You find a composition, **A*B**, in a Cayley table like this one, by locating **A** along the left edge and **B** along the top edge.




| * | I | $R_{90}$ | $R_{180}$ | $R_{270}$ | H | V | D | D′ |
|---|---|---|---|---|---|---|---|---|
| **I** | I | $R_{90}$ | $R_{180}$ | $R_{270}$ | H | V | D | D′ |
| **$R_{90}$** | $R_{90}$ | $R_{180}$ | $R_{270}$ | I | D′ | D | H | V |
| **$R_{180}$** | $R_{180}$ | $R_{270}$ | I | $R_{90}$ | V | H | D′ | D |
| **$R_{270}$** | $R_{270}$ | I | $R_{90}$ | $R_{180}$ | D | D′ | V | H |
| **H** | H | D | V | D′ | I | $R_{180}$ | $R_{90}$ | $R_{270}$ |
| **V** | V | D′ | H | D | $R_{180}$ | I | $R_{270}$ | $R_{90}$ |
| **D** | D | V | D′ | H | $R_{270}$ | $R_{90}$ | I | $R_{180}$ |
| **D′** | D′ | H | D | V | $R_{90}$ | $R_{270}$ | $R_{180}$ | I |

**A Cayley table for the symmetries of a square**

For example, the yellow cell in the table says that **H*$R_{90}$** = **D**. That is, first performing **$R_{90}$** and then performing **H** results in the same final position as **D,** as illustrated below:



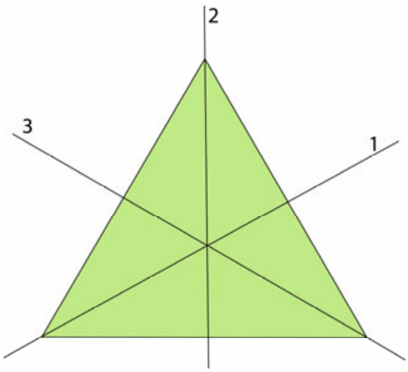

**$H*R_{90}$ = D because "performing H after performing $R_{90}$" yields the same ending position as performing D.**

Notice that the **H** reflection line does not rotate along with **$R_{90}$** because **H** is a fixed rigid motion of the plane.

This table provides a wealth of information about the symmetries of the square. What patterns do you see? What can you learn from this table? Here is one important observation: **$H*R_{90}$** = **D**, while **$R_{90}$***H** = **D′**. Order matters! "Performing **$R_{90}$** and then **H"** IS DIFFERENT THAN "performing **H** and then **$R_{90}$**."

Next, let us try an equilateral triangle, which has the following six symmetries:

**I** = the rotation by zero (the identity)
**$R_{120}$** = rotation by 120°
**$R_{240}$** = rotation by 240°
**$F_1$** = flip over line 1
**$F_2$** = flip over line 2
**$F_3$** = flip over line 3.




**The triangle's reflection lines**

Fill in the table below showing the composition of any two of these symmetries. For this task, you will need a cardboard triangle with vertices labeled A, B, and C. To get you started, the yellow entry means that **$R_{120}$*$F_1$ = $F_2$.** When you are done, you will have a Cayley table for the triangle. What patterns do you see? How is it similar to the square's Cayley table? How is it different?

| * | I | $R_{120}$ | $R_{240}$ | $F_1$ | $F_2$ | $F_3$ |
|---|---|---|---|---|---|---|
| I | | | | | | |
| $R_{120}$ | | | | $F_2$ | | |
| $R_{240}$ | | | | | | |
| $F_1$ | | | | | | |
| $F_2$ | | | | | | |
| $F_3$ | | | | | | |

A Cayley table for the symmetries of a triangle

## Symmetry Groups

Composing symmetries has a lot in common with adding and multiplying numbers. To investigate this similarity, let us first review the familiar algebraic properties of multiplication and addition of numbers.

- ➔ The order in which a pair of numbers are added (or multiplied) does not affect the result. This is called the <u>commutative</u> property. In symbols:

  $$A+B = B+A \quad \text{and } A \times B = B \times A$$

- ➔ The order in which a pair of additions (or a pair of multiplications) is performed does not affect the result. This is called the <u>associative</u> property. In symbols:

  $$(A+B)+C = A+(B+C) \text{ and } (A \times B) \times C = A \times (B \times C)$$

- ➔ Adding 0 to a number has no effect. Multiplying a number by 1 has no effect. In symbols:

  $$A + 0 = 0 + A = A \quad \text{and} \quad 1 \times A = A \times 1 = A$$

  We call 0 and 1 the <u>identities</u>; 0 is the additive identity and 1 is the multiplicative identity.

- ➔ For any number A, the sum of A and –A equals 0, the additive identity. For any nonzero number A, the product of

A and 1/A equals 1, the multiplicative identity. We call –A the additive inverse of A, and 1/A the multiplicative inverse of A.

Composition of symmetries shares all of these features except for the commutative property. Do you see the similarities? Imagine three friends:

(1) Adam likes to add numbers. Adam has a special number, 0, which has no effect when he adds it to other numbers. That is his identity. He can find any number's "inverse" (the thing that adds to it to give his identity). For example, the inverse of 35 is –35.
(2) Michelle likes to multiply numbers. Michelle has a special number, 1, which has no effect when she multiplies it by other numbers. That is her identity. She can find any number's "inverse" (the thing that multiplies by it to give her identity). For example, the inverse of 35 is 1/35.
(3) Chris likes to compose symmetries of the square. Chris has a special symmetry, **I**, which has no effect when he composes it with other symmetries. That is his identity. He can find any symmetry's "inverse" (the thing that composes with it to give his identity). For example, the inverse of $\mathbf{R_{90}}$ is $\mathbf{R_{270}}$.

What do these three stories have in common, aside from corny alliteration? Each of these three friends is working with a system that mathematicians refer to as a "group." In common English, the word "group" has a very general meaning, but in mathematics, it has a very specific technical meaning. Its meaning is meant to capture the commonalities between the systems studied by Adam, Michelle, and Chris in the above stories.

**DEFINITION:** A group is a set (denoted G) with an algebraic operation (denoted •) that satisfies the following properties:

(1) G has an "identity" (denoted **I**) which has no effect on other members; that is, A•**I** = A and **I**•A = A for all members, A, of G.

(2) Each member, A, of G has an inverse in G (usually denoted $A^{-1}$), which combines with it in either order to give the identity: $A{\bullet}A^{-1} = \mathbf{I}$ and $A^{-1}{\bullet}A = \mathbf{I}$.

(3) The associative property holds: (A•B)•C = A•(B•C) for all triples A,B,C of members of G.

If the commutative property also holds (which means A•B = B•A for all pairs A,B of members of G), then we call G a commutative group; otherwise we call G a non-commutative group.

Think of "•" as a generic symbol, which Adam replaces with "+,"Michelle replaces with "×" and Chris replaces with "*." In our previous discussion of these three friends, we were too vague about what the word "number" means. We now clarify this point.

**EXAMPLE:** Adam likes the infinite set of all integers, **Z** = {… –3, –2, –1, 0, 1, 2, 3, …}, which is a commutative group under the operation of addition. The identity is 0, and the inverse of A is –A.

**EXAMPLE:** Michelle likes the set of all non-zero real numbers, which is a commutative group under the operation of multiplication. The identity is 1, and the inverse of A is 1/A.

**NON-EXAMPLE:** The integers, **Z** = {… –3, –2, –1, 0, 1, 2, 3, …}, is NOT a group under the operation of division. Most integer divisions, like 5/7, do not result in integers. A valid algebraic operation for a group G must be a method of combining each pair of members of G to give an answer that is a member of G.

**NON-EXAMPLE:** The integers, **Z** = {… –3, –2, –1, 0, 1, 2, 3, …}, is NOT a group under the operation of multiplication. Although 1 is the identity, the number 3 does not have an inverse in **Z** (because 1/3 is not an integer).

Chris likes the eight symmetries of a square, which form a group. There is nothing special about a square; the collection of symmetries of ANY object forms a group!

**THEOREM:** The collection of symmetries of any object in the plane forms a group under the operation of composition "*."

This group is called the symmetry group of that object. Its identity is the rigid motion that we previously called "the identity." The inverse of a symmetry, A, is denoted $A^{-1}$; it is the rigid motion that undoes what A does.

One step of proving this theorem is checking that the associative property is valid within symmetry groups. If A, B, and C are rigid motions, it is difficult to visualize why (A*B)*C is the same rigid motion as A*(B*C). The trick is to imagine what happens to any single point of the plane, visualized as a single drop of paint on the infinite glass wall. Both (A*B)*C and A*(B*C) effect a single point by first performing C then B and then A. They have the same effect on each single point, so they must be the same rigid motion.

Another step is verifying that every symmetry of an object has an inverse. More generally, every rigid motion of the plane has an inverse – a rigid motion that undoes it.

**THEOREM:** Every rigid motion of the plane has an inverse.

Although we will not give a formal proof of this fact, it should seem believable. No matter what rigid motion I apply to the plane, you can always move it back into its starting position. If I rotate

27° clockwise, then you can rotate 27° counterclockwise. If I flip and then translate, then you can translate back and then flip back. If the original motion is video recorded, then its inverse can be visualized by imagining running the video backwards.

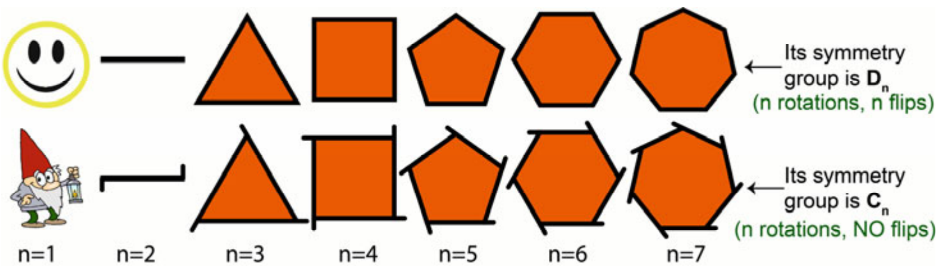
The symmetry groups of the oriented and non-oriented regular polygons are so important, we have special names and symbols for them:

**DEFINITION:** Suppose that $n \geq 2$.

The symmetry group of a regular $n$-gon is denoted as $\mathbf{D_n}$ and is called the $n$th dihedral group.

The symmetry group of an *oriented* regular $n$-gon is denoted as $\mathbf{C_n}$ and is called the $n$th cyclic group.

Recall (from Exercise #3 in Chap. 1) that $\mathbf{D_n}$ has $n$ rotations and $n$ flips, while $\mathbf{C_n}$ has $n$ rotations and no flips (the identity counts as a rotation). This pattern can be extended to $n = 1$ by defining $\mathbf{D_1}$ and $\mathbf{C_1}$ to mean the symmetry groups of the smiley face and gnome, respectively, as shown in the illustration below.




Representative objects whose symmetry groups are $D_n$ and $C_n$

## The Power of Inverses

The existence of inverses is surprisingly useful. The remainder of this chapter is devoted to some of its powerful consequences. For example, did you notice that the square's

Cayley table looked like a Sudoku board – all eight symmetries appeared in each row and column. There is nothing special about a square; we will prove that any Cayley table has this property.

**THE SUDOKU THEOREM:** If an object has a finite symmetry group, then every symmetry appears exactly once in each row and each column of its Cayley table.

**PROOF:** To show that an arbitrary symmetry, called **A**, must appear in the column of the symmetry **B**, we must locate a symmetry which, when performed after **B**, yields **A**. The correct choice is ($\mathbf{A}*\mathbf{B}^{-1}$). This works because:

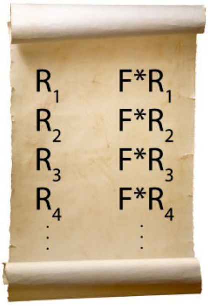
$$(\mathbf{A}*\mathbf{B}^{-1})*\mathbf{B} = \mathbf{A}*(\mathbf{B}^{-1}*\mathbf{B}) = \mathbf{A}*\mathbf{I} = \mathbf{A}.$$

There are as many symmetries as positions in the column; since each symmetry appears there, each must appear exactly once. The claim about rows is proven similarly……………………………□

The idea of this proof actually shows that the Cayley table of any finite group (not just a symmetry group) has the Sudoku property. Here is another beautiful consequence of the existence of inverses. It is not a coincidence that a regular polygon has equal numbers of rotations and flips. If an object has any improper symmetries at all, then it has equal numbers of proper and improper symmetries. In other words:

**THE ALL-OR-HALF THEOREM:** If an object has a finite symmetry group, then either all or half of its symmetries are proper.

**PROOF:** Suppose that NOT all of the object's symmetries are proper. Choose one of the improper symmetries and call it **F** (for Flip). On the left column of a parchment of ancient scroll, list all of the proper symmetries: $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{R}_3$, etc. On the right column, show the result of composing these proper symmetries with **F** (as




illustrated). Your left column is now a list of all of the proper symmetries. Your right column is a list of all of the improper symmetries. The left and right columns have the same sizes; thus, there are equal numbers of proper and improper symmetries. Why are the symmetries in the right column all improper? Because an improper symmetry composed with a proper symmetry is always improper. Why does every improper symmetry appear somewhere in the right column, with no repetitions? Because every symmetry appears exactly once in **F**'s row of the Cayley table..................................................................................□

Our final application of the existence of inverses has to do with objects that lack symmetry.

**DEFINITION:** An object is called <u>asymmetric</u> if it has no symmetries other than the identity.

A haphazard doodle will almost certainly be asymmetric. The next theorem says that asymmetric objects are very useful as rigid motion detectors:

**RIGID MOTION DETECTOR THEOREM:** If an object is asymmetric, then any rigid motion of the plane is uniquely determined by knowing the object's appearance after that motion is applied.

To understand this theorem, let us first think about why it is NOT true for a symmetric object, like a square. Suppose you close your eyes and then reopen them to discover that the square has moved 3 in. to the right. From this, you can NOT tell what motion I performed while your eyes were closed. I might have translated 3 in. to the right or I might have rotated 90° and then translated 3 in. to the right. You have no way of knowing. The rotation part is undetectable because it is a symmetry of the square. The following proof shows that, if the square in this story is replaced by an asymmetric object, then there could not be any undetectable part.

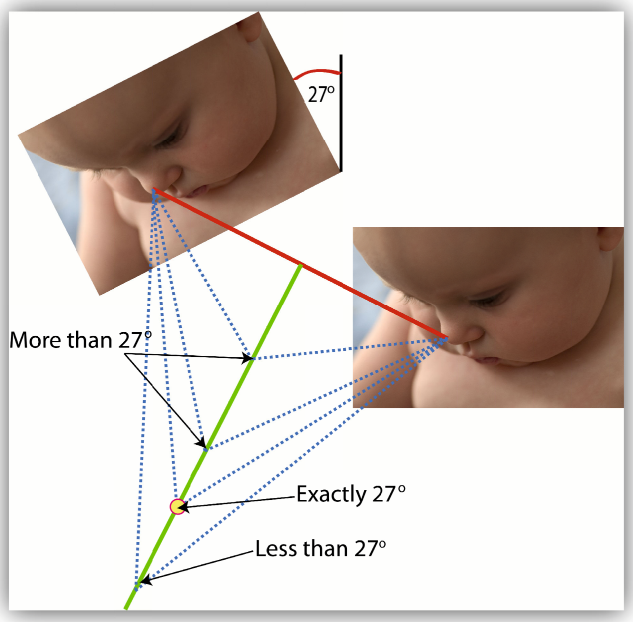
**PROOF:** Suppose that two different rigid motions, called A and B, have exactly the same effect on an asymmetric object, so that when you open your eyes you cannot tell whether A or B were performed. But if A and B were genuinely different, then $A^{-1}*B$ would be a symmetry of the object that is different from the identity, which is impossible because the object is asymmetric...............□

The previous gnome image is asymmetric; that is why it visually distinguished the symmetries of the square.

## An Improved Classification of Plane Rigid Motions

Since the baby photo shown here is asymmetric, the previous theorem implies that there is ONLY ONE rigid motion that moves it from its bottom-right position to its top-left position. It seems to have been rotated counterclockwise by 27° (you could measure this angle with a protractor) and then translated up and left. In fact, you could achieve that same repositioning of the photo by only doing one thing: rotating by 27° about the correctly chosen point. How do you find this point?

Here's how. First, chose a distinguished point on the photo, like the nose, and draw a red line connecting the two noses. Next, draw a green perpendicular bisector. The point we seek lies on this green line, and it is chosen so that the




illustrated blue hinge measures exactly 27°. Rotating the plane by 27° about this special point will achieve the illustrated repositioning of the baby photo, since it moves the nose to the proper place and it tilts the photo the proper amount! If you do not own a protractor, there are alternative methods you could use to solve this problem. For example, you could select several distinguished points (nose, earlobe, and eyelash), and locate the intersection of all of the corresponding green lines. But with our protractor method, we can easily explain why the method works. As you slide down the green line, there is clearly a unique point where the blue hinge angle will change from too large to too small. We now have the key idea for proving the first part of the following important theorem:

**CLASSIFICATION OF PLANE RIGID MOTIONS (VERSION 2):**

Every <u>proper</u> rigid motion of the plane is a translation or a rotation.

Every <u>improper</u> rigid motion of the plane is a reflection or a glide reflection.

**PROOF OF FIRST CLAIM** : Assuming version 1 of the classification theorem (in Chap. 1), all we must prove is this: A rotation followed by a translation is the same as a single rotation. We will use the baby photo (or any other asymmetric image) as our rigid motion detector. A rotation followed by a translation has the same effect on the baby photo as a single rotation whose center is found using the previously explained protractor method. Since it has the same effect on the baby photo, it must be the same rigid motion.......................................................................□

We will not prove the second claim (about improper rigid motions), but we encourage you to think about how to prove it.

## Exercises

(1) If you have not yet done so, fill in the Cayley table for the triangle in this chapter.

(2) Describe any patterns you see in the Cayley tables for $\mathbf{D_3}$ and $\mathbf{D_4}$, which were constructed in the chapter.

(3) Construct a Cayley table for $\mathbf{D_2}$. Is $\mathbf{D_2}$ a commutative group?

(4) Construct a Cayley table for $\mathbf{C_n}$ for each $n = 2, 3, 4, 5$. Describe any patterns and similarities you see. How is the Cayley table for $\mathbf{C_n}$ related to the Cayley table for $\mathbf{D_n}$?

(5) Construct a Cayley table for $\mathbf{D_5}$. *HINT: Construct a cardboard pentagon with labeled vertices. Use your pentagon to fill in some of the Cayley table, and then save time by using the Sudoku property to fill in the rest.*

(6) Find the inverse of each of the symmetries in the Cayley tables you constructed in Exercises (3), (4), and (5).

(7) Does the G border pattern (an infinite strip of Gs) have a commutative symmetry group?

…. G G G G G G G G G G G G …..

What about the C border pattern? What about Z? What about Y?

(8) Which symmetry of the square must be performed *after* **H** to yield $\mathbf{R_{270}}$? Which must be performed *before* **D** to yield **V**?

(9) In $\mathbf{D_3}$ = the symmetry group of a triangle, solve the following equation for X:

$$\mathbf{R_{120}} * X = \mathbf{F_1}$$

Solve this in two ways. First, scan down $\mathbf{R_{120}}$'s row of the Cayley table until you find $\mathbf{F_1}$ – the answer is the column in which you find it. Second, left compose each side of the equation with the inverse of $\mathbf{R_{120}}$. In any group, do you think you can always solve an equation of the form A*X = B for X? How?

(10) In $\mathbf{D_3}$, solve the following equation for X:

$$X * \mathbf{R_{120}} = \mathbf{F_1}$$

Solve this in two ways. First, scan down $\mathbf{R_{120}}$'s column of the Cayley table until you find $\mathbf{F_1}$ – the answer is the row in which you find it. Second, right compose each side of the equation with the inverse of $\mathbf{R_{120}}$. In any group, do you think you can always solve an equation of the form X*A = B for X? How?

(11) Decide whether the following statements are true or false, and discuss:

(1) Each cyclic group is commutative
(2) Each dihedral group is noncommutative

(12) Prove that every rotation symmetry of a border pattern is by 0 or 180°.

*HINT: Suppose that some rotation (called **R**) by a different angle were a symmetry. If **T** is a translation symmetry, explain why* $\boldsymbol{R}*\boldsymbol{T}*\boldsymbol{R}^{-1}$ *equals a translation symmetry that is NOT parallel to **T**? What can you conclude?*

(13) Do the four improper symmetries of a square alone form a group under composition?

(14) Is the set of even integers a group under addition? What about the set of odd integers?

(15) If an object has infinitely many symmetries, explain why it must have either zero or infinitely many improper symmetries.

(16) What does the scroll in the proof of the All-or-Half Theorem look like if the object is a square and the improper symmetry you choose is **H**?

(17) What type of symmetry results if you first perform a reflection over one line, and next perform a reflection over a second line? *(HINT: either the two lines are parallel or they intersect in a point. Consider each of these possibilities separately.)*

(18) Can a bounded object have only reflection symmetries (no rotations other than the identity)? If so, then how many different reflection symmetries could it have? Explain. *(HINT: if it had more than one reflection symmetry, what could you conclude using your solution to the previous exercise?)*

(19) Physicists still do not fully understand why galaxies have spiral shapes. What is the symmetry group of the spiral approximated by each spiral galaxy pictured below?

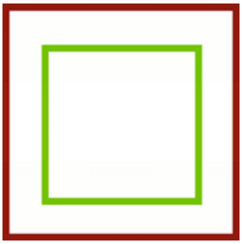
**NGC 1300 photo by European Southern Observatory**

**Whirlpool galaxy photo by NASA and ESA**

# 3. Isomorphism

This chapter is about the things that matter and the things that do not matter. Consider this. In the last chapter, you learned that $\mathbf{D_4}$ means the symmetry group of a square. Did you respond by asking: "Which square? Where is it centered? How big is it? Is it upright or tilted? Is it green or purple?" Later, when you studied the symmetry group of an infinite strip of Gs, did you ask: "How tall are the Gs? How far apart are they spaced? What color are they?" You probably did NOT ask these questions because you intuitively sensed that their answers do not matter. In exactly what sense do these things not matter? To focus on what does matter, we will need a more precise way of understanding exactly what does not matter. That is the purpose of this chapter. That is the purpose of an isomorphism.

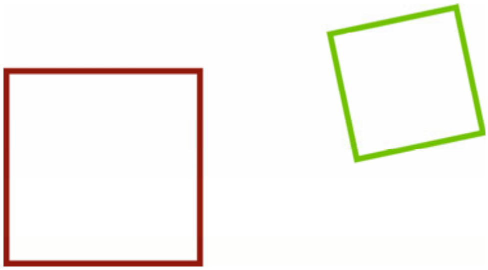
First recall that it is possible for two different objects to have exactly the same symmetry groups. In the illustration on the right, every symmetry of the larger red square is also a symmetry of the smaller green square, and vice-versa, so the symmetry groups of the red square and of the green square are literally identical.


Two squares with the same symmetries

But what happens if the two squares have different centers and are tilted at different angles, as pictured below? If a rigid motion of the plane is performed that is a symmetry of the red square, it will move the green square to another location; thus, the symmetry groups of the red and green squares are not

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_3,

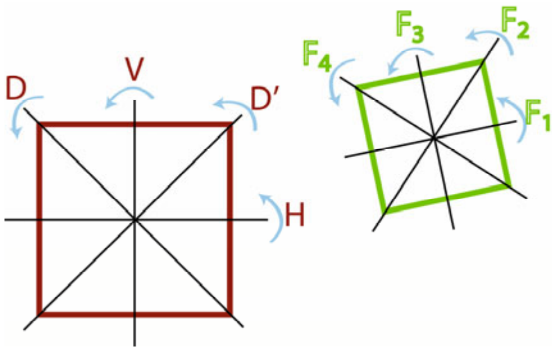
literally identical. Nevertheless, we intuitively sense that the two symmetry groups are still essentially the same – they have the same essential algebraic structure within their Cayley tables. We will call a pair of groups "isomorphic" if they are essentially the same. The goal of this chapter is to formulate a precise way to express this concept.


Two squares with different symmetries

## What Is an Isomorphism?

Let us think carefully about the relationship between the symmetry groups of the red and green squares which are illustrated above.

Rene is studying the symmetry group of the red square. It has the following eight symmetries: {**I**, $\mathbf{R_{90}}$, $\mathbf{R_{180}}$, $\mathbf{R_{270}}$, **H**, **V**, **D**, **D′**}. Her Cayley table looks exactly like the Cayley table in Chap. 2 (typeset in red). Meanwhile, Gretchen is studying the symmetry group of the green square.




Her square has the following eight symmetries: {**I**, $\mathbf{R_{90}}$, $\mathbf{R_{180}}$, $\mathbf{R_{270}}$, $\mathbf{F_1}$, $\mathbf{F_2}$, $\mathbf{F_3}$, $\mathbf{F_4}$}. Since her reflection lines are not quite horizontal, vertical or diagonal, she chose different symbols for naming them. Even her rotations differ from Rene's rotations. For example, Rene's $\mathbf{R_{90}}$ is a rotation about the center of the red square, while Gretchen's $\mathbf{R_{90}}$ is a rotation about the center of the green square – these are different rigid motions of the plane.

Nevertheless, Gretchen’s Cayley table is essentially the same as Rene’s in the following precise sense. Starting with Rene’s red Cayley table, if I ask my word processor to translate red symbols into green symbols using the following symbol-replacement dictionary:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| I | $R_{90}$ | $R_{180}$ | $R_{270}$ | H | D′ | V | D |
| ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ |
| I | $R_{90}$ | $R_{180}$ | $R_{270}$ | $F_1$ | $F_2$ | $F_3$ | $F_4$ |

I will end up with a valid Cayley table for Gretchen’s green group.

| * | I | $R_{90}$ | $R_{180}$ | $R_{270}$ | H | V | D | D’ |
|---|---|---|---|---|---|---|---|---|
| I | I | $R_{90}$ | $R_{180}$ | $R_{270}$ | H | V | D | D’ |
| $R_{90}$ | $R_{90}$ | $R_{180}$ | $R_{270}$ | I | D’ | D | H | V |
| $R_{180}$ | $R_{180}$ | $R_{270}$ | I | $R_{90}$ | V | H | D’ | D |
| $R_{270}$ | $R_{270}$ | I | $R_{90}$ | $R_{180}$ | D | D’ | V | H |
| H | H | D | V | D’ | I | $R_{180}$ | $R_{90}$ | $R_{270}$ |
| V | V | D’ | H | D | $R_{180}$ | I | $R_{270}$ | $R_{90}$ |
| D | D | V | D’ | H | $R_{270}$ | $R_{90}$ | I | $R_{180}$ |
| D’ | D’ | H | D | V | $R_{90}$ | $R_{270}$ | $R_{180}$ | I |

| * | I | $R_{90}$ | $R_{180}$ | $R_{270}$ | $F_1$ | $F_3$ | $F_4$ | $F_2$ |
|---|---|---|---|---|---|---|---|---|
| I | I | $R_{90}$ | $R_{180}$ | $R_{270}$ | $F_1$ | $F_3$ | $F_4$ | $F_2$ |
| $R_{90}$ | $R_{90}$ | $R_{180}$ | $R_{270}$ | I | $F_2$ | $F_4$ | $F_1$ | $F_3$ |
| $R_{180}$ | $R_{180}$ | $R_{270}$ | I | $R_{90}$ | $F_3$ | $F_1$ | $F_2$ | $F_4$ |
| $R_{270}$ | $R_{270}$ | I | $R_{90}$ | $R_{180}$ | $F_4$ | $F_2$ | $F_3$ | $F_1$ |
| $F_1$ | $F_1$ | $F_4$ | $F_3$ | $F_2$ | I | $R_{180}$ | $R_{90}$ | $R_{270}$ |
| $F_3$ | $F_3$ | $F_2$ | $F_1$ | $F_4$ | $R_{180}$ | I | $R_{270}$ | $R_{90}$ |
| $F_4$ | $F_4$ | $F_3$ | $F_2$ | $F_1$ | $R_{270}$ | $R_{90}$ | I | $R_{180}$ |
| $F_2$ | $F_2$ | $F_1$ | $F_4$ | $F_3$ | $R_{90}$ | $R_{270}$ | $R_{180}$ | I |

Do you see why the resultant green table is a *valid* Cayley table for Gretchen’s green group (which means that each of its 64 cells is filled in correctly)? If you start checking cells, you will quickly see why things are working out. Our symbol-replacement dictionary was very carefully constructed. The reflection lines listed counterclockwise around the red square were matched with the reflection lines listed counterclockwise around the green square. Said differently, this dictionary is induced by the rigid motion that slides and tilts the red square on top of the green square, matching up the two sets of reflection lines.

In summary, each cell in the green table is filled in correctly, which really means this: Our dictionary translates every true red equation into a true green equation! For example the yellow highlighted cell of the red table represents “**$H*R_{90}$ = D**” which is a true red equation. This equation is symbol-by-symbol translated into “**$F_1*R_{90}$ = $F_4$**” which is a true green equation, represented by the yellow highlighted cell of the green table.

The symmetry groups of the two squares are “isomorphic.” The dictionary between their members is an “isomorphism.” Here is the precise definition:

**DEFINITION:** An isomorphism between two groups means a one-to-one matching (dictionary) between their members that translates each true equation in one group into a true equation in the other. This is the same as saying that it translates an entire Cayley table for one group into a valid Cayley table for the other.

We say two groups are isomorphic if there exists an isomorphism between them.

When two groups are isomorphic, we think of them as essentially the same; they have the same algebraic structure within their Cayley tables. Imagine an observer who does not know or care what the names of the group members represent (symmetries of a red square or of a green square or whatever). This observer would study the two Cayley tables and discover exactly the same patterns. From this observer’s perspective, the two groups would look like a single group represented in two different notational systems.

An isomorphism is analogous to a language dictionary. For example, the square’s Cayley table in Chap. 2 would look different if this textbook was translated into Swahili. Cosmetic differences would arise because the Swahili word for "Rotation" does not begin with "R," nor does the word for "Horizontal" begin with "H." But these irrelevant notational differences do not matter – Swahili readers will learn the same things as English readers. It is a single group represented in two different notational systems. The dictionary between the English and Swahili notational systems is like an isomorphism. It translates true equations into true equations. For example, when the Swahili symbol for **H** is composed with the Swahili symbol for $\mathbf{R_{90}}$ it had better equal the Swahili symbol for **D**; otherwise the translator needs to be fired.

## Isomorphism Examples

**Example: The Star and The Moth**

Moth photo by Ken Tapp

The swirly star above has exactly two symmetries: the identity and a rotation by 180° about its center (like $\mathbf{C_2}$). The moth also has two symmetries: the identity and a vertical flip (like $\mathbf{D_1}$). Here are their Cayley tables:

| *star* | I | $R_{180}$ |
|---|---|---|
| I | I | $R_{180}$ |
| $R_{180}$ | $R_{180}$ | I |

| *moth* | I | V |
|---|---|---|
| I | I | V |
| V | V | I |

The symmetry group of the star is isomorphic to the symmetry group of the moth. The isomorphism is the following one-to-one matching between their members: **I** ↔ **I** $\mathbf{R_{180}}$ ↔ **V**.

This isomorphism translates each star symmetry into a moth symmetry. It deserves to be called an isomorphism because it translates the star's entire Cayley table into the moth's entire Cayley table. It does not matter that a flip is geometrically different from a rotation. All that matters is the underlying algebraic patterns in the Cayley tables. An observer who did not know the meaning of the symbols $\mathbf{R_{180}}$ or **V** would look at the two Cayley tables and describe them in the same way: they both have exactly one non-identity member which is its own inverse. Thus, $\mathbf{C_2}$ is isomorphic to $\mathbf{D_1}$.

**Example: George and Peter**

George is studying the symmetry group of a border pattern constructed from an infinite strip of green Gs:

…. G G G G G G G G G G G G G G ….

The subsequent Gs are 1 cm apart. He calls his group G. Its members are: G = {…,$T_{-3}$, $T_{-2}$, $T_{-1}$, $T_0$, $T_1$, $T_2$, $T_3$,…}, where $T_n$ denotes the translation by $n$ centimeters to the right (if $n$ is positive) or to the left (if $n$ is negative).

Meanwhile, Peter is studying the symmetry group of a border pattern constructed from an infinite strip of purple Ps:

… P P P P P P P P P …

The subsequent Ps are 2 cm apart. He calls his group P. Its members are P = {…,$T_{-6}$, $T_{-4}$, $T_{-2}$, $T_0$, $T_2$, $T_4$, $T_6$,…}. George and Peter chat over coffee and come to suspect that their groups might be isomorphic. The alleged isomorphism they construct is the following one-to-one matching:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| … | $T_{-4}$ | $T_{-3}$ | $T_{-2}$ | $T_{-1}$ | $T_0$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ | … |
| | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | |
| … | $T_{-8}$ | $T_{-6}$ | $T_{-4}$ | $T_{-2}$ | $T_0$ | $T_2$ | $T_4$ | $T_6$ | $T_8$ | … |

The pattern doubles each green subscript, matching $T_3$ with $T_6$ for example. Since $T_3$ and $T_6$ each translates its border pattern three letters to the right, George and Peter believe that this is an isomorphism. They feel that counting letters matters more than counting centimeters, and that color does not matter a bit.

To check that their matching is really an isomorphism, they must verify that each true green equation becomes a true purple equation. For example, "$T_5*T_8 = T_{13}$" is a true green equation. If we double all subscripts, we get "$T_{10}*T_{16} = T_{26}$", which is in fact a true purple equation. There is nothing special about 5 and 8; for any integers *m* and *n*, their matching gives the following:

$$\begin{array}{ccccc} T_m & * & T_n & = & T_{m+n} \\ \downarrow & & \downarrow & & \downarrow \\ T_{2m} & * & T_{2n} & = & T_{2m+2n}. \end{array}$$

The translated purple equation is true. This verifies that each true green equation is translated into a true purple equation, and therefore that their matching is really an isomorphism!

**Example: George and The Integers**

When our friend George (from the previous example) composes his symmetries, he is really just adding centimeters. For example, $T_5*T_8 = T_{13}$ because 5+8 = 13. He is performing integer addition in disguise. He is suddenly struck by a revelation: his group is isomorphic to the additive group of integers,

$$\mathbf{Z} = \{\ldots,-3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

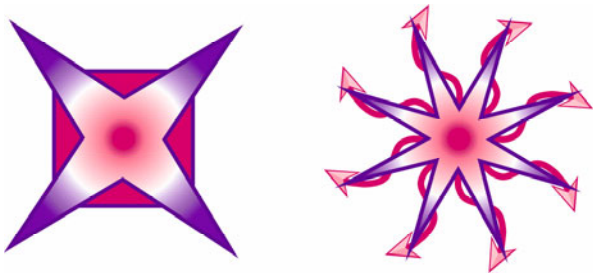
The isomorphism that he constructs between G and **Z** simply matches $T_n$ with *n*. For example, it matches the symmetry $T_7$ with the integer 7. As evidence that this is really an isomorphism, George observes that the true equation "$T_5*T_8 = T_{13}$" in G is translated into "5 + 8 = 13," which is a true equation in **Z**. Notice that the translated equation has "+" instead of "*" because + is the algebraic operation in **Z**. As before, there is nothing special about 5 and 8; every true equation in G is translated into a true equation in **Z**:

$$\begin{array}{ccccc} T_m & * & T_n & = & T_{m+n} \\ \downarrow & & \downarrow & & \downarrow \\ m & + & n & = & m+n. \end{array}$$

Since G is a symmetry group of a border pattern and **Z** is not a symmetry group at all (it is just a bunch of integers), George worries that he is comparing apples and oranges when he claims that G and **Z** are isomorphic. In fact, he is allowed to make this comparison because G and **Z** have in common that they are both groups.

**NON-EXAMPLE:** The dihedral group $\mathbf{D_4}$ is NOT isomorphic to the cyclic group $\mathbf{C_{10}}$ because they have different sizes. $\mathbf{D_4}$ has eight members, while $\mathbf{C_{10}}$ has ten members. Isomorphic groups always have the same size, since an isomorphism is a one-to-one matching between their members.

**NON-EXAMPLE:** The dihedral group $\mathbf{D_4}$ is NOT isomorphic to the cyclic group $\mathbf{C_8}$. Even though they both have eight members, $\mathbf{C_8}$ is commutative while $\mathbf{D_4}$ is non-commutative. A commutative group could never be isomorphic to a non-commutative group. Think about why. Thus, although each object pictured below has eight symmetries, there is an essential difference between their symmetry groups (which are $\mathbf{D_4}$ and $\mathbf{C_8}$, respectively). These two objects are symmetric in different ways.



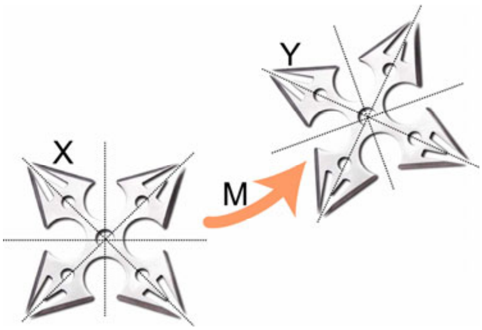
Each star has 8 symmetries, but their symmetry groups are not isomorphic.

## Rigid Equivalence

In the previous chapter, we defined $\mathbf{D}_n$ to mean the symmetry group of a regular *n*-gon. We did not bother to specify irrelevant details like where the *n*-gon was centered or how it was tilted. These details do not matter because moving or rotating an object (or transforming it by any rigid motion) does not really change its symmetry group. Here is the precise way to express this:

**THEOREM:** Performing a rigid motion does not essentially change an object's symmetry group. More precisely, its symmetry groups before and after the object is transformed by the rigid motion are isomorphic to each other.

**PROOF:** Let us call the rigid motion M, the original object X, and the transformed object Y, as illustrated below. It is visually apparent that M matches every symmetry of X with a symmetry of Y. For example, M moves each reflection line of X to a reflection line of Y, and M moves the rotation center of X to the rotation center of Y.




Here is a clever way to more precisely specify the manner in which M induces a matching between the symmetries of X and Y. Every symmetry, A, of X gets matched with $M*A*M^{-1}$. Do you see why $M*A*M^{-1}$ is a symmetry of Y? First $M^{-1}$ is performed (which moves Y to X), then A is performed (which moves X to X because A is a symmetry of X), and then M is performed (which moves X back to Y). The net result is a symmetry of Y.

To verify that this matching is an isomorphism, we must check that it translates true equations into true equations. Suppose A*B = C is a true equation in the symmetry group of X. This equation translates into $(M*A*M^{-1})*(M*B*M^{-1}) = M*C*M^{-1}$, which is a true equation because its left side simplifies to its right side as follows:

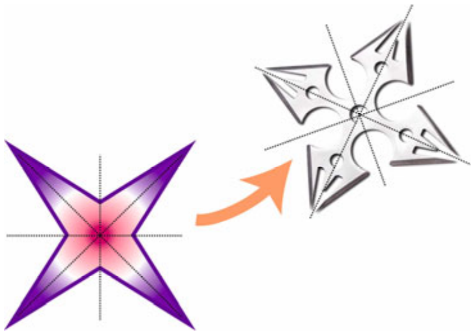
$$\text{LEFT} = M*A*M^{-1}*M*B*M^{-1} = M*A*I*B*M^{-1} = M*A*B*M^{-1}$$
$$= M*C*M^{-1} = \text{RIGHT}.$$

Think on your own about why this matching is one-to-one……....□

Since the symmetry group of an object is unaffected by rigid motions, it is useful to define the following:

**DEFINITION:** Two objects in the plane are called rigidly equivalent if there exists a rigid motion of the plane which, when applied to the first object, repositions it so that afterwards the two objects have exactly the same symmetries.

The two stars pictured are rigidly equivalent because the purple one can be tilted and slid on top of the silver one so that, after this repositioning, the two objects have symmetry groups which are not just isomorphic but are literally identical. Every symmetry of the silver star is a symmetry of the repositioned purple star, and every symmetry of the repositioned purple star is a symmetry of the silver star.



These two stars are rigidly equivalent

**THEOREM:** If two objects are rigidly equivalent, then their symmetry groups are isomorphic.

**PROOF:** The symmetry group of the first object is isomorphic to the symmetry group of the repositioned first object, which is identical to the symmetry group of the second object……….……………□

Let us reconsider the examples from the previous section in terms of this new notion of rigid equivalence:

**[The red & green squares]** These two squares are rigidly equivalent, which is the real reason that their symmetry groups turned out to be isomorphic.

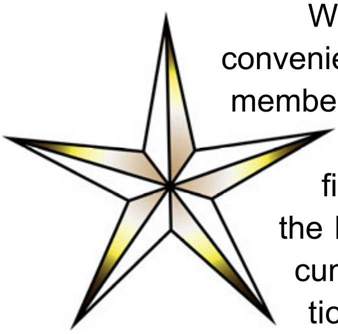
**[The star & the moth]** The symmetry group of the star is isomorphic to the symmetry group of the moth, even though the star is NOT rigidly equivalent to the moth. A rigid motion cannot change the fact that the star has a rotation, while the moth has a flip.

**[George & Peter]** George's G border pattern is NOT rigidly equivalent to Peter's P border pattern. A rigid motion cannot change the fact that their smallest translation lengths (1 and 2 cm respectively) are different. If the G border pattern was enlarged by a scaling factor of 2, then the two border patterns would become rigidly equivalent.

These examples help us to understand the interrelated concepts of rigid equivalence and isomorphism. It is all about what matters and what does not matter. In the next chapter, we will answer this difficult question: What are all of the possible ways in which (1) a bounded objects, (2) a border pattern, and (3) a wallpaper patterns can be symmetric? But before answering this question, we must decide what it means. When do we consider two objects to be symmetric in the same way? Should this mean that the objects are rigidly equivalent? Or should it mean that the objects have isomorphic symmetry groups? Or maybe it should mean that one object is rigidly equivalent to a rescaling (enlarging or shrinking) of the other object. All of these possibilities are reasonable, and in some cases, these different possibilities lead to different answers to the question.

That is the reason this chapter is important. Our main challenge in this book is to understand the different ways in which objects can be symmetric. The concepts of isomorphism and rigid equivalence provide us with a language for more precisely saying what exactly this challenge entails.

## A Better Notation for the Cyclic Groups



We end this chapter by describing a convenient and simple new notation system for the members of a cyclic group. The 5th cyclic group $\mathbf{C_5} = \{\mathbf{I}, \mathbf{R_{72}}, \mathbf{R_{144}}, \mathbf{R_{216}}, \mathbf{R_{288}}\}$ contains the five rotations of the oriented star pictured on the left. Keeping track of all of these angles is cumbersome. Instead, it is common convention to name the members of $\mathbf{C_5}$ like this: $\mathbf{C_5} = \{0, 1, 2, 3, 4\}$. Think of "3" as representing "$\mathbf{R_{216}}$" which is a rotation of the star by 3 counterclockwise "turns." We are using the word "turn" here to mean the smallest possible rotation angle, 72°, which moves each star point to its counterclockwise neighbor. Thus, we are counting turns instead of angles. The Cayley table for $\mathbf{C_5}$ looks particularly simple when translated into this concise new notational system:

| * | **I** | $\mathbf{R_{72}}$ | $\mathbf{R_{144}}$ | $\mathbf{R_{216}}$ | $\mathbf{R_{288}}$ |
|---|---|---|---|---|---|
| **I** | **I** | $\mathbf{R_{72}}$ | $\mathbf{R_{144}}$ | $\mathbf{R_{216}}$ | $\mathbf{R_{288}}$ |
| $\mathbf{R_{72}}$ | $\mathbf{R_{72}}$ | $\mathbf{R_{144}}$ | $\mathbf{R_{216}}$ | $\mathbf{R_{288}}$ | **I** |
| $\mathbf{R_{144}}$ | $\mathbf{R_{144}}$ | $\mathbf{R_{216}}$ | $\mathbf{R_{288}}$ | **I** | $\mathbf{R_{72}}$ |
| $\mathbf{R_{216}}$ | $\mathbf{R_{216}}$ | $\mathbf{R_{288}}$ | **I** | $\mathbf{R_{72}}$ | $\mathbf{R_{144}}$ |
| $\mathbf{R_{288}}$ | $\mathbf{R_{288}}$ | **I** | $\mathbf{R_{72}}$ | $\mathbf{R_{144}}$ | $\mathbf{R_{216}}$ |

| + | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 4 |
| 1 | 1 | 2 | 3 | 4 | 0 |
| 2 | 2 | 3 | 4 | 0 | 1 |
| 3 | 3 | 4 | 0 | 1 | 2 |
| 4 | 4 | 0 | 1 | 2 | 3 |

It is also common convention to use "+" to denote the algebraic operation in a cyclic group. The yellow highlighted cells above represent the equation that we previously wrote as "$\mathbf{R}_{216}$* $\mathbf{R}_{288}$ = $\mathbf{R}_{144}$" and which we now write as "3 + 4 = 2 (in $\mathbf{C}_5$)".

This notational change encourages you to think of the algebraic operation in $\mathbf{C}_5$ as encoding a kind of "addition with wrap-around." As you count 4 past 3 on your fingers, remember that when you reach 5, you must wrap back around to zero, so your count goes: 4, 0, 1, 2.

In our future study of $\mathbf{C}_n$, we will always use this abbreviated notation system:

$$\mathbf{C}_n = \{0,1,2,\ldots,n-1\},$$

and we will always write "+" instead of "*." For example, we will write:

$$5 + 7 = 2 \text{ (in } \mathbf{C}_{10})$$

to indicate that, when you rotate a 10-gon by 7 turns and then by 5 more turns, you have done 2 turns more than going all of the way around, so the result is the same as rotating just two turns (here the word "turn" means a rotation by 360/10 = 36°). Alternatively, just add on your fingers, wrapping back to zero when you reach ten, so your count of 7 past 5 goes "6, 7, 8, 9, 0, 1, 2." Alternatively, just add 5 + 7 = 12, and then subtract 10 to get 2. It is all the same thing.

Here is another illuminating example:

$$9 + 7 = 4 \text{ (in } \mathbf{C}_{12})$$

In $\mathbf{C}_{12}$, addition-with-wrap-around works like a clock: 9 h after 7 o'clock, it will be 4 o'clock. However, this analogy is imperfect because, when computing in $\mathbf{C}_{12}$, only the numbers between 0 and 11 are used. The analogy between $\mathbf{C}_{12}$ and clock-arithmetic only becomes airtight if we all agree to henceforth say “0” instead of “12” when reporting the time of day, so that midnight and noon are called “zero o’clock.” We hope that you will not mind.

This new notational system for $\mathbf{C}_n$ is not exactly an isomorphism between two groups because there was only ever one group in the story. Nevertheless, it fits well into the chapter about isomorphism. When a pair of groups is isomorphic, it is often best to think of them as a single group described in two different notation systems.

## Exercises

(1) Fill in each blank:

5 + 7 = ___ (in $\mathbf{C}_9$), 6 + ___ = 2 (in $\mathbf{C}_{10}$), 80 + 35 = ___ (in $\mathbf{C}_{100}$).

(2) Construct a complete Cayley table for $\mathbf{C}_7$.

(3) Draw several different objects whose symmetry groups are all isomorphic to $\mathbf{D}_5$.

(4) Draw several different objects whose symmetry groups are all isomorphic to $\mathbf{C}_6$.

(5) Could enlarging or shrinking an object ever essentially change its symmetry group?

(6) Some copy machines let you separately control the vertical and horizontal scaling factors when you enlarge or shrink a picture. Could a vertical-only shrinking ever essentially change an object’s symmetry group?

(7) Take the 26 capital English letters, and sort them into piles according to whether their symmetry groups are isomorphic. Within each pile, sub-sort them according to whether the letters are rigidly equivalent.

(8) Explain why the symmetry group of the Z border pattern is NOT isomorphic to the symmetry group of the G border pattern.

(9) Verify that the following one-to-one matching is NOT an isomorphism between the red and green squares in this chapter:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **I** | $\mathbf{R_{90}}$ | $\mathbf{R_{180}}$ | $\mathbf{R_{270}}$ | **H** | **D′** | **V** | **D** |
| ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ |
| **I** | $\mathbf{F_1}$ | $\mathbf{R_{180}}$ | $\mathbf{R_{270}}$ | $\mathbf{R_{90}}$ | $\mathbf{F_2}$ | $\mathbf{F_3}$ | $\mathbf{F_4}$ |

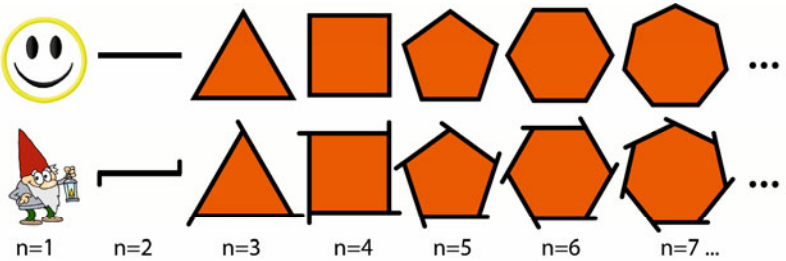
Is a random haphazard matching of the red and green symbols likely to be an isomorphism?

(10) Make Cayley tables for $\mathbf{D_2}$ and $\mathbf{C_4}$ and explain why these groups are NOT isomorphic.

(11) In the following list of groups, identify a pair which is isomorphic, and explain why there are no other such pairs:

$$\mathbf{C_1, D_1, C_2, D_2, C_3, D_3, C_4, D_4, C_5, D_5, C_6, D_6, \ldots}$$

(12) In the following list of objects, identify a pair which has isomorphic symmetry groups and explain why there are no other such pairs. Is any pair rigidly equivalent?



(13) Draw two border patterns which have isomorphic symmetry groups but yet can NOT be enlarged or shrunk to make them rigidly equivalent.

(14) Must any two asymmetric objects be rigidly equivalent?

# 4. The Classification Theorems

Leonardo Da Vinci's self-portrait

Historically, the concept of symmetry evolved slowly from a vague idea to a precise notion as scientists and mathematicians sought to study the symmetry of their world using ever more precise language and methods. The classification theorems in this chapter represent some of the pinnacles of this historic journey. These theorems provide complete classifications of the possible ways in which each kind of planar object we have studied (bounded objects, border patterns, and wallpaper patterns) can be symmetric!

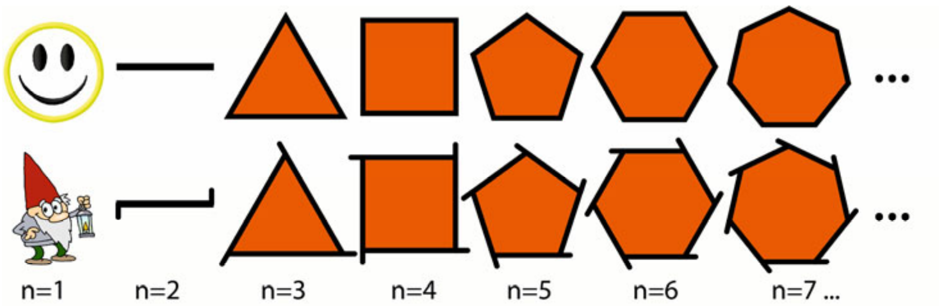
## Bounded Objects

The description of all possible symmetry groups of bounded objects is usually attributed to Leonardo Da Vinci (1452–1519).

**DA VINCI'S THEOREM:** The symmetry group of any bounded object in the plane is either infinite or is isomorphic to a dihedral or cyclic group.

Thus, if the object's symmetry group is finite, then Da Vinci's Theorem guarantees that it is isomorphic to the symmetry group of one of these objects pictured below:

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_4, 




Representatives of all the ways in which a bounded object can be symmetric

Actually, a stronger "rigid" version of this theorem is true: Any bounded object in the plane with a finite symmetry group is rigidly equivalent to one of the objects pictured above! Of course, you must imagine that the list of pictures above does not stop at $n = 7$, but goes on indefinitely. This rigid version is the one that Leonardo Da Vinci understood, since he predated the development of group theory.

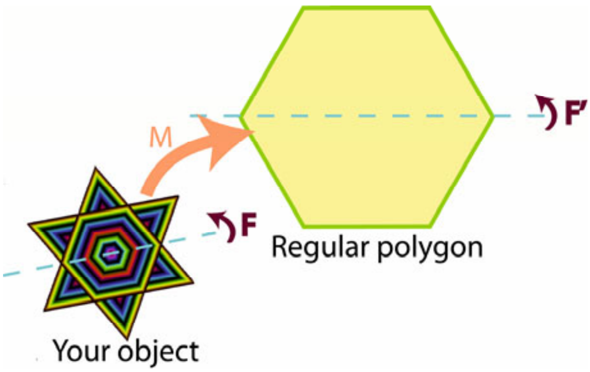
Notice that Da Vinci's Theorem tells you nothing about a bounded object whose symmetry group is infinite. You might guess that such an object must be rigidly equivalent to a circle, but this is not quite true. There are some nuances involved which we will not be equipped to discuss until a later chapter.

**PROOF OF RIGID VERSION OF DA VINCI'S THEOREM**: Imagine that you have a bounded object with a finite symmetry group. The Center Point Theorem tells us that all of its rotations have the same center. In fact, we claim that they must all be multiples of the smallest-angle rotation. Let us use an example to think about why. If the smallest rotation symmetry was $\mathbf{R}_{10}$, then $\mathbf{R}_{20}$, $\mathbf{R}_{30}$, $\mathbf{R}_{40}$,…,$\mathbf{R}_{350}$ would also be symmetries. Something else, like $\mathbf{R}_{37}$, could not also be a symmetry, because that would make $\mathbf{R}_{37}*(\mathbf{R}_{30})^{-1} = \mathbf{R}_7$ be a rotation symmetry smaller than $\mathbf{R}_{10}$. Do you see how this example leads you to the general idea needed to justify the claim?

Say your object has *n* rotation symmetries. Since these rotations are all multiples of the smallest, they have the same angles as the rotations of a regular *n*-gon. If your object has no flips, then it is rigidly equivalent to an oriented regular *n*-gon via any rigid motion that matches up their center points.

If your object has any flips, then choose one flip and call it **F**. Composing **F** with each of the *n* rotations yields a list of *n* different flips. By the All-Or-Half Theorem, your object has no flips other than these *n*.

Next draw a regular *n*-gon, choose one of its flips, and call it **F′**. All of the *n*-gon's flips are obtained by composing its *n* rotations with **F′**. Now move your object on top of the regular *n*-gon which you drew, using a rigid motion, M, that matches their center points and aligns the **F**-flip line with the **F′**-flip line, as illustrated on the right. After this repositioning, your object will have exactly the same symmetries as the regular *n*-gon, namely, the same rotations about their now-common center point, and the compositions of these rotations with their now-common chosen flip. Thus, your object is rigidly equivalent to the regular *n*-gon. In particular, your object's symmetry group is isomorphic to that of the regular *n*-gon ....................................................................□




**COROLLARY:** If a bounded object has exactly *n* rotations and zero flips, then its symmetry group is isomorphic to $\mathbf{C}_n$. If a bounded object has exactly *n* rotations and *n* flips, then its symmetry group is isomorphic to $\mathbf{D}_n$.

It is interesting that there are no other possibilities. Before reading Da Vinci's Theorem, you might have imagined that it was possible to sketch a picture whose *n* rotations and *n* flips fit together into a Cayley table that is essentially different from $\mathbf{D_n}$.

## Border Patterns

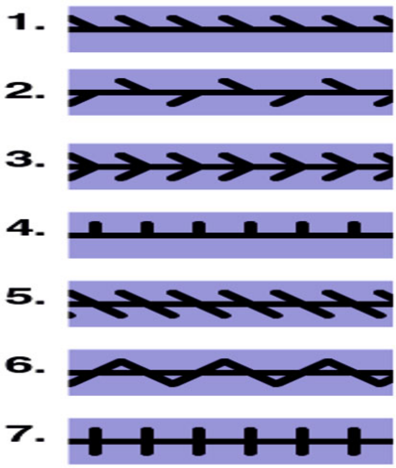
Images by Becky F from Flickr.com

Photo by Horia Varlan from Flickr.com

The border patterns which you encounter in art and architecture exhibit a seemingly infinite variety of artistic motifs. However, from a mathematical point of view, there are only seven different ways in which a border pattern can be symmetric!

**THE CLASSIFICATION OF BORDER PATTERNS:** Any border pattern is rigidly equivalent to a rescaling of one of the seven model border patterns illustrated below (provided it has a smallest non-identity translation).

The requirement that the border pattern has a smallest non-identity translation excludes patterns like an infinite horizontal line, which differs from the seven model patterns in that it can be translated any distance right or left, no matter how small. Some texts include the smallest translation requirement into their definition of a border pattern.




The 7 model border patterns: Image by AndrewKepert on Wikipedia.org

A "rescaling" (enlarging or shrinking) is just needed to ensure that the border pattern has the same smallest translation length as the model pattern to which it is being compared.

Rescaling a border pattern does not essentially change its symmetry group. Thus, the group version of the classification theorem says this: The symmetry group of any border pattern is isomorphic to the symmetry group of one of the seven model border patterns (provided it has a smallest translation). However, among the seven model patterns, some of the symmetry groups are isomorphic to each other. After removing the redundant ones, the list of seven shrinks to a list of four. Not all isomorphisms are explained by rigid motions.

If you have a border pattern, how do you know which of the seven model patterns yours is rigidly equivalent to? All you must do is fill out the following "identification card":

**Q1** – Does it have any horizontal reflection symmetry?
**Q2** – Does it have any vertical reflection symmetry?
**Q3** – Does it have any 180° rotation symmetry?
**Q4** – Does it have any glide reflection symmetry?

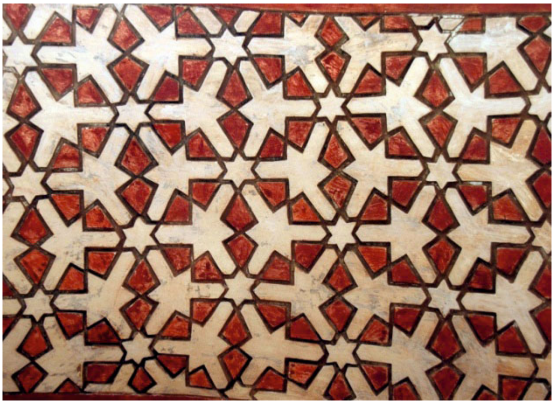
Border pattern identification card

Your pattern will be rigidly equivalent to the model pattern that has the same yes/no responses to all four questions on this identification card. The term "horizontal" here really means the direction of your border pattern's translations, and the term “vertical” really means the direction perpendicular to your border pattern’s translations.

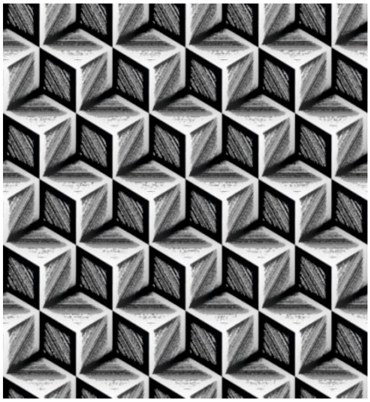
In fact, the steps involved in proving the classification theorem are: (1) Prove that a border pattern cannot have any types of symmetries other than translations and those types mentioned on the identification card, (2) prove that two border patterns with the same identification cards must be rigidly equivalent to rescaling of each other, and (3) decide which combinations of yes/no responses are identification cards of actual border patterns.

## Wallpaper Patterns

Like border patterns, wallpaper patterns exhibit endless artistic variety, but yet we will soon learn that there are only 17 ways in which a wallpaper pattern can be symmetric.



**Qubbah Ba'adiyim in Marrakesh**
***photo by amerune, Flickr.com***

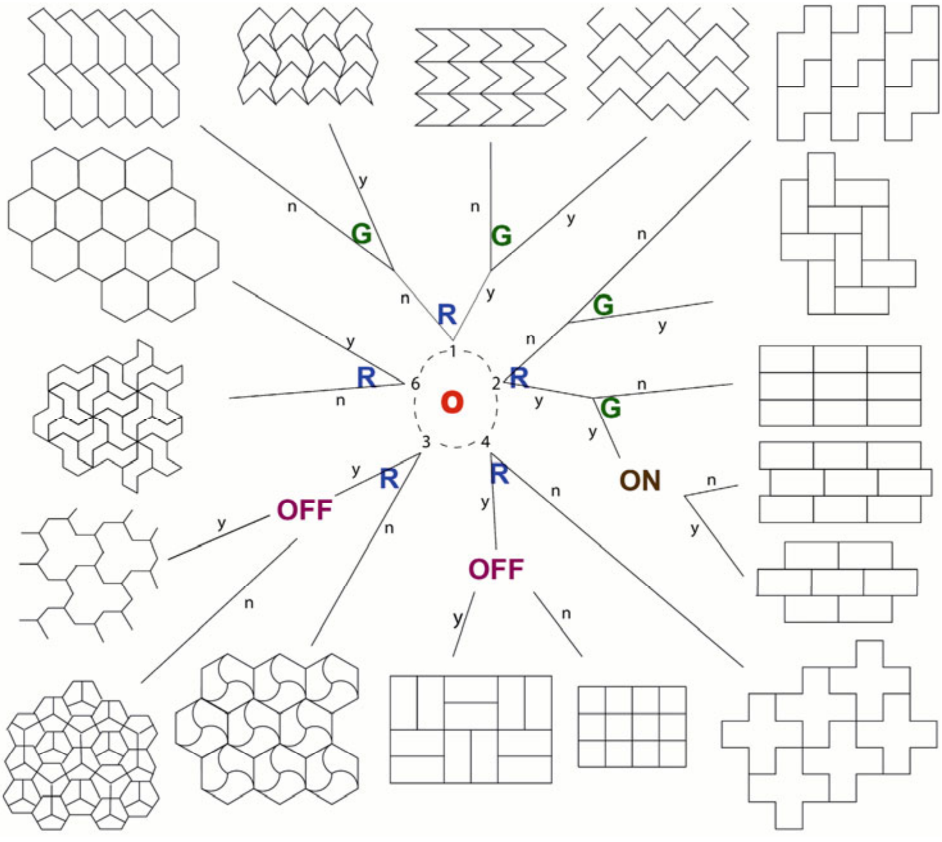


**WoodCut QBert Block Texture**
***by Patrick Hoesly, Flickr.com***

Just like border patterns, wallpaper patterns are classified according to their answers to the questions on an identification card. The first question is this: What is the maximal order of a rotation symmetry of the pattern? This means the number of times that the rotation must be repeatedly performed to return to the starting position. For example a 90° rotation has order 360/90 = 4, while a 60° rotation has order 360/60 = 6. Do you see the pattern? It is quite surprising that the only possible orders for the rotation symmetries of wallpaper patterns are 1, 2, 3, 4 and 6, as we will soon see.

Another identification card question asks whether there are any glide reflection symmetries that are indecomposable. This means that the translation and reflection out of which the glide reflection is built are not themselves individually symmetries of the pattern. Below are the identification card and the flow chart by which the 17 wallpaper patterns are classified:

**O** – What is the maximum Order of a rotation symmetry?
**R** – Does it have any Reflection symmetries?
**G** – Does it have an indecomposable Glide reflection symmetries?
**ON** – Does it have any rotations centered ON reflection lines?
**OFF** – Does it have any rotations centered OFF reflection lines?




The 17 model wallpaper patterns: diagram by Brian Sanderson,http://www.warwick.ac.uk/~maaac/

If you have a wallpaper pattern, then your pattern will be "symmetric in the same way" as one of the 17 model patterns; namely, the one whose identification card places it in the same flow chart position as your pattern. To state a precise theorem, we must decide what "symmetric in the same way" means. One possibility is to focus on symmetry groups:

**THE CLASSIFICATION OF WALLPAPER PATTERNS:** The symmetry group of any wallpaper pattern is isomorphic to the symmetry group of one of the 17 model patterns (provided it has a smallest non-identity translation).

However, this theorem is not quite optimal because, among the 17 model patterns, certain pairs have isomorphic symmetry groups, so a list of fewer than 17 model patterns would have sufficed. There is a rigid version of this theorem, which genuinely requires all 17 model patterns, but it is a bit too technical to fully describe in this book. Here is the rough idea. To make an arbitrary wallpaper pattern become rigidly equivalent to one of the 17 model patterns, you must alter it by something called a <u>linear transformation</u>, which is slightly more general than a rescaling because it includes things like vertical-only-shrinkings/enlargings and <u>shears</u> (which are illustrated in Exercise 9). We are not quite equipped to give a precise definition of a linear transformation. That is why we only stated the group version of the classification theorem. This theorem is difficult to prove; a complete proof did not appear until 1891.

The symmetry groups of the 17 model patterns are often called "wallpaper groups." Chemists call them "plane crystallographic groups" because they represent the possible configurations into which two dimensional crystal structures can form. M.C. Escher incorporated many of these patterns into his paintings. They also occur throughout nature, for example in honeycombs.

## Summary

In this chapter, we did something remarkable. We described all possible ways in which (1) a bounded object, (2) a border pattern, and (3) a wallpaper pattern can be symmetric. For each of these three categories, we found a list of models. For bounded objects, our models were the oriented and non-oriented regular polygons, plus a gnome and a happy face. For border patterns, our list contained exactly seven models. For wallpaper patterns, our list contained exactly 17 models.

Any object in the category is “symmetric in the same way” as one of the models. What exactly does it mean for two objects to be “symmetric in the same way?” It depends on the category. For bounded objects, the best meaning is that the objects are rigidly equivalent. For border patterns, the best meaning is that one pattern is rigidly equivalent to a rescaling of the other. For wallpaper patterns, the best meaning is that one object is rigidly equivalent to the result of performing a linear transformation to the other, but since linear transformations are difficult to define, we settled for the next-best meaning: that the patterns have isomorphic symmetry groups.

This chapter represents yet another exhibition of the importance of precise language in mathematics. Before you can answer a question, you must understand exactly what the question means. Since the first chapter of this book, we have had a growing desired to understand the ways in which objects can be symmetric. One of the most difficult challenges turned out to be the clarification of precisely what this means.

## Exercises

(1) For each of the seven model border patterns, answer the four identification card questions. If Q1 and Q2 both answer yes for a pattern, why must Q3 and Q4 also answer yes?

(2) For each capital English letter, decide which of the 7 model border patterns is formed by repeating it indefinitely along a line.

(۞3) Of the 7 model border patterns, find two which have isomorphic symmetry groups.

(۞4) Which of the 7 model border patterns have commutative symmetry groups?

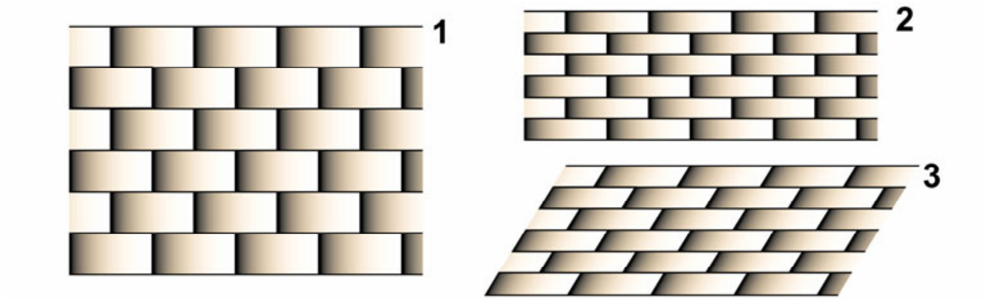
(5) How many of the 7 model border patterns are oriented?

(6) How many of the 17 model wallpaper patterns are oriented?

(7) Of the 17 model wallpaper patterns, identify at least one which has a commutative symmetry group and at least one which does not. Explain why.

(۞8) Prove that a border pattern could not have any reflection symmetries other than horizontal and vertical.

(9) Brick patterns 2 and 3 below were obtained from 1 by applying linear transformations; namely, a vertical-only-shrinking for 2 and a “shear” for 3. Imagine that all three patterns are extended indefinitely to form wallpaper patterns. Decide which of the three are “symmetric in the same way” as each other, according to several meanings for this term: (A) they have isomorphic symmetry groups, (B) they are rigidly equivalent, (C) they are rigidly equivalent to rescaling of each other, (D) they have identical identification cards.




(10) Classify each of these borders patterns as type 1-7.

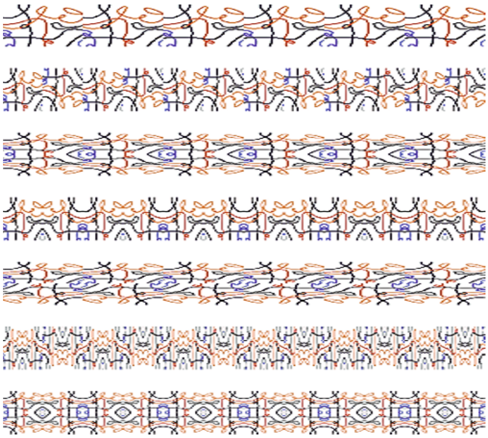
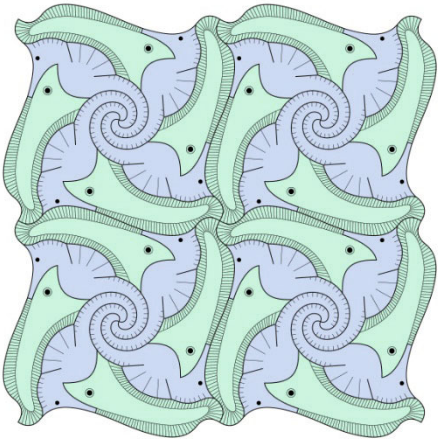
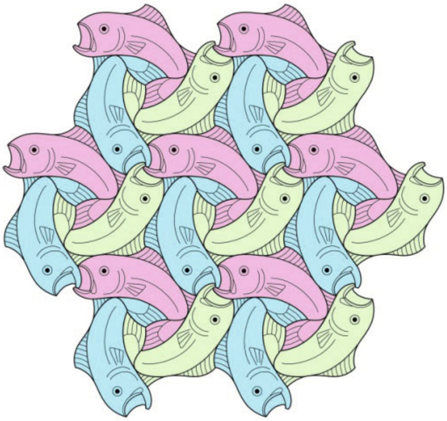


Image by User:Cyp from wikipedia.org

(11) Perform a web image search for “M.C. Escher symmetry.” Many of Escher’s paintings are wallpaper patterns (if you imagine them infinitely extended). How many of the 17 model wallpaper patterns can you find represented in his paintings?

(12) Identify the model wallpaper pattern that matches with the paintings *Seahorses and Eels* and *Three Fishes* by Robert Fathauer. How would the answer change if the fish all had the same color?





(13) Identify the model wallpaper patterns that match with *Qubbah Ba'adiyim* and *WoodCut QBert Block Texture,* illustrated in the chapter.

(14) Identify the model border patterns that match with the border patterns by Beck F and Horia Varlan illustrated in the chapter.

# 5. Subgroups and Product Groups

In this chapter, we learn how to find small groups inside of large groups, and then how to build large groups out of small groups. The point is to help you better understand symmetry groups. If you can recognize an object's symmetry group as having been built out of smaller groups, then this realization might help you to much more clearly understand its underlying algebraic structure.

## Subgroups

First, we learn how to find small groups inside of large groups. For example, inside $\mathbf{D_4}$ = {**I**, $\mathbf{R_{90}}$, $\mathbf{R_{180}}$, $\mathbf{R_{270}}$, **H**, **V**, **D**, **D′**}, let us separately consider the rotations R = {**I**, $\mathbf{R_{90}}$, $\mathbf{R_{180}}$, $\mathbf{R_{270}}$} and the flips F = {**H**, **V**, **D**, **D′**}. Let us build red-only and green-only tables, like this:

| * | I | $R_{90}$ | $R_{180}$ | $R_{270}$ |
|---|---|---|---|---|
| I | I | $R_{90}$ | $R_{180}$ | $R_{270}$ |
| $R_{90}$ | $R_{90}$ | $R_{180}$ | $R_{270}$ | I |
| $R_{180}$ | $R_{180}$ | $R_{270}$ | I | $R_{90}$ |
| $R_{270}$ | $R_{270}$ | I | $R_{90}$ | $R_{180}$ |

| * | H | V | D | D' |
|---|---|---|---|---|
| H | I | $R_{180}$ | $R_{90}$ | $R_{270}$ |
| V | $R_{180}$ | I | $R_{270}$ | $R_{90}$ |
| D | $R_{270}$ | $R_{90}$ | I | $R_{180}$ |
| D' | $R_{90}$ | $R_{270}$ | $R_{180}$ | I |

The red table is the Cayley table for R = {**I**, $\mathbf{R_{90}}$, $\mathbf{R_{180}}$, $\mathbf{R_{270}}$}, which is a self-contained group that happens to be isomorphic to $\mathbf{C_4}$. The green table does not look like a Cayley table at all, due to the fact that F = {**H**, **V**, **D**, **D′**} is not a self-contained group. After studying the green table, you can identify several reasons why F is not a group. First, there is no identity. Second, the cells of the green table are filled with symbols that are not in F. If F were

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_5,

really a group, then each cell of its Cayley table would be filled with something in F – there could be no foreign symbols.

To summarize: the rotations, R, form a special type of collection, called a "subgroup," because it is a self-contained group in its own right. The special properties of R, by virtue of which we know it is a self-contained group are: (1) the identity is a rotation, (2) the inverse of every rotation is a rotation, and (3) the composition of every pair of rotations is a rotation (so that the red Cayley table has no foreign symbols). Look back at the definition of "group" and think about why these three properties are exactly what is needed to verify that R is a group. The flips, F, do NOT form a self-contained group because (1) the identity is not a flip and (2) the composition of a pair of flips is not a flip.

**DEFINITION:** Suppose that G is a group. A collection, H, of G's members is called a subgroup if it forms a self-contained group on its own, which means that it satisfies all three of these requirements:

(Identity) H must include the identity of G.
(Products) If A and B are in H, then A • B must be in H.
(Inverses) The inverse of anything in H must be in H.

We have seen that the rotations form a subgroup of $\mathbf{D_4}$, but the flips do not. Here are some more examples.

**EXAMPLE:** The collection K = {**I, $\mathbf{R_{180}}$, H, V**} is a subgroup of $\mathbf{D_4}$. Why? First, it contains the identity, **I**. Second, each thing in K is its own inverse. Third, the product (composition) of any two things in K always lies in K, because every cell of the table at the right is filled in with one of the four things in K (no foreign symbols). In fact,

| * | I | H | V | $R_{180}$ |
|---|---|---|---|---|
| **I** | **I** | **H** | **V** | $\mathbf{R_{180}}$ |
| **H** | **H** | **I** | $\mathbf{R_{180}}$ | **V** |
| **V** | **V** | $\mathbf{R_{180}}$ | **I** | **H** |
| $\mathbf{R_{180}}$ | $\mathbf{R_{180}}$ | **V** | **H** | **I** |

K looks like $\mathbf{D_2}$ = the symmetry group of the 2-gon. Think of K here as the symmetries of the square which are also symmetries of the 2-gon drawn as a horizontal middle stripe.

**EXAMPLE:** The symmetry group of the W border pattern,

… W W W W W W W W W W W W W W W W W …

includes translations and also vertical flips. The translations alone form a subgroup. We will denote these translations like this: T = $\{\ldots,T_{-3},T_{-2},T_{-1},T_0,T_1,T_2,T_3,\ldots\}$. Why is it a subgroup? First, it includes the identity, $T_0$. Second, the inverse of every translation is a translation. For example, the inverse of $T_7$ is $T_{-7}$. Third, the composition of any pair of translations is a translation. For example, the composition of $T_7$ and $T_8$ is $T_{15}$.

**NON-EXAMPLE:** In the symmetry group of the W border pattern, the flips alone do NOT form a subgroup. The problems are: (1) the identity is not a flip, and (2) the composition of a pair of flips is not a flip – it is a translation.

**EXAMPLE:** In the additive group of integers, **Z** = {…,–3, –2, –1, 0, 1, 2, 3, …}, the evens, **E** = {…, –6, –4, –2, 0, 2, 4, 6, …} form a subgroup. Why is **E** a subgroup of **Z**? First, the identity, 0, is even. Second, the inverse of every even number is even. For example, the inverse of 6 is –6. Third, the sum of any two even numbers is even (remember that adding is the algebraic operation in this group). Think about why the odd numbers do NOT form a subgroup of **Z**.

**EXAMPLE:** In the cyclic group $\mathbf{C_{10}}$ = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, the even numbers, **E** = {0, 2, 4, 6, 8} form a subgroup. Why is **E** a subgroup of $\mathbf{C_{10}}$?

First, the identity, 0, is even. Second, the inverse of every even number is even: the inverse of 2 is 8, the inverse of 4 is 6, and the inverse of 0 is 0. Third, the "sum with wrap-around" of any two even numbers is even, because the Cayley table for **E** (shown here) is entirely populated by even numbers. Can you explain why **E** is isomorphic to $\mathbf{C_5}$?

| + | 0 | 2 | 4 | 6 | 8 |
|---|---|---|---|---|---|
| 0 | 0 | 2 | 4 | 6 | 8 |
| 2 | 2 | 4 | 6 | 8 | 0 |
| 4 | 4 | 6 | 8 | 0 | 2 |
| 6 | 6 | 8 | 0 | 2 | 4 |
| 8 | 8 | 0 | 2 | 4 | 6 |

**NON-EXAMPLE:** In the cyclic group $\mathbf{C_9}$ = {0, 1, 2, 3, 4, 5, 6, 7, 8}, the even numbers, **E** = { 0, 2, 4, 6, 8 } do NOT form a subgroup. The first problem is that the inverse of an even number is NOT always even in this group. For example, the inverse of 2 is 7 and the inverse of 4 is 5. Second, the sum-with-wrap-around of two even numbers is not always even, like 6 + 8 = 5 (in $\mathbf{C_9}$).

## Generating Subgroups

Suppose you have a group, G, and you wish to find a subgroup of it. You could list all of G's members, and then circle the ones you are putting into your subgroup. But how do you do the circling so that you end up with a valid subgroup? Here is one way. First, circle the identity, I, because it is in every subgroup. Next choose some other member, A, to circle. Now you are forced to circle the inverse of A. You are also forced to circle the product of A with itself any number of times, and the product of the inverse of A with itself any number of times. After you have done all of this, you will have a valid subgroup. It contains A plus everything else that is forced to be there by the three requirements of being a subgroup. We call this subgroup <A> because it is "generated" by A.

**DEFINITION:** If G is a group, and A is a member of G, then,

<A> = {…, $A^{-1}$•$A^{-1}$•$A^{-1}$, $A^{-1}$•$A^{-1}$, $A^{-1}$, **I**, A, A•A, A•A•A,…}

is called the subgroup of G generated by A. It contains A and $A^{-1}$ combined with themselves any number of times.

**EXAMPLE:** In **Z** = {…, –3, –2, –1, 0, 1, 2, 3, …}, let us take A = 3. The inverse of 3 is –3. The symbol "•" must be replaced with "+," which is the algebraic operation in **Z**. For example, the expression "A•A•A" becomes "3+3+3 = 9". Similarly, "$A^{-1}$•$A^{-1}$•$A^{-1}$•$A^{-1}$" becomes "–3+–3+–3+–3 = –12" and so on. We conclude that <3> = { …, –12, –9, –6, –3, 0, 3, 6, 9, 12, …} = the multiples of 3.

**EXAMPLE:** In $\mathbf{D_4}$, let us take A = $\mathbf{R_{90}}$. The inverse of A is $\mathbf{R_{270}}$. The symbol "•" must be replaced with "*" because composition is the algebraic operation in $\mathbf{D_4}$. We have:

A = $\mathbf{R_{90}}$, A•A = $\mathbf{R_{180}}$, A•A•A = $\mathbf{R_{270}}$, A•A•A•A = **I**, …

(we can stop here because the pattern repeats).

$A^{-1}$ = $\mathbf{R_{270}}$, $A^{-1}$•$A^{-1}$ = $\mathbf{R_{180}}$, $A^{-1}$•$A^{-1}$•$A^{-1}$ = $\mathbf{R_{90}}$, $A^{-1}$•$A^{-1}$•$A^{-1}$•$A^{-1}$ = **I,...**

(we can stop here because the pattern repeats). In summary,

<A> = {**I,** $\mathbf{R_{90}}$, $\mathbf{R_{180}}$, $\mathbf{R_{270}}$}.

This makes sense – if you perform a 90° rotation forward or backwards as many times as you like, you generate all of the rotations of the square.

In the last example, when we composed A with itself repeatedly, we eventually arrived at the identity, after which the pattern began repeating. Furthermore, the things obtained by composing $A^{-1}$ with itself repeatedly were redundant to the things obtained by composing A with itself repeatedly. In fact, this is exactly what always happens in any finite group.

**THEOREM:** If G is a *finite* group, and A is a member of G, then

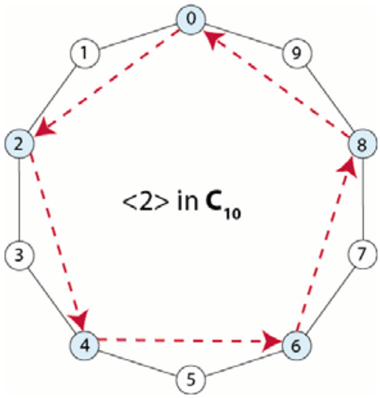
<A> = { I, A, A•A, A•A•A, A•A•A•A, …}

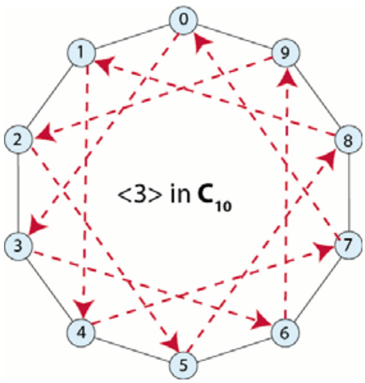
(this list starts repeating as soon as one of these expressions equals I, and not before).

This theorem tells us how to quickly find the members of <A>. Just write the identity, then write A, then A•A, and so on until you arrive at the identity again, at which point you are done. For example, in $\mathbf{C}_{10}$ = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9},

<2> = {0, 2, 4, 6, 8} (the evens)
<3> = {0, 3, 6, 9, 2, 5, 8, 1, 4, 7} (all of $\mathbf{C}_{10}$)
<4> = {0, 4, 8, 2, 6} (the evens – the same as <2>)
<5> = {0, 5}.

There is a nice way to visualize these subgroups of $\mathbf{C}_{10}$. Imagine 10 friends (numbered 0–9) standing in a circle. They pass a ball amongst themselves. Friend number 0 starts with the ball and passes it to her right. The subgroup <2> is the collection of friends who touch the ball if everyone passes 2 to their right. The subgroup <3> is the collection of friends who touch the ball if everyone passes 3 to their right and so on.



Try drawing your own such circle diagram to illustrate <4> and <5> in $\mathbf{C}_{10}$. Perhaps these diagrams help you see why the **C**s are called *cyclic* groups.

We previously defined the "order" of a rotation symmetry, so that $\mathbf{R}_{360/n}$ has order *n*. More generally, we now define "order" of any member of any group:

**DEFINITION:** If A is a member of a finite group, then the order of A is the size of the subgroup <A>.

The "size" of a group or subgroup just means its number of members. Thus, the order of A is the smallest number of copies of A which must be combined to get the identity. For example, the order of the member 2 of the group $\mathbf{C}_{10}$ equals 5 because 2+2+2+2+2 = 0, and no smaller list of 2s sum to the identity. The order of the member $\mathbf{R}_{90}$ of the group $\mathbf{D}_4$ equals 4 because a 90° rotation must be done 4 times in succession to return a square to its starting position.

## Product Groups

Next, we consider an important way of putting together a pair of groups, $G_1$ and $G_2$, to build a single new group, which is denoted $G_1 \times G_2$ and is called the product of $G_1$ and $G_2$. The members of this new group are all of the possible ways of pairing together a member of the first group with a member of the second group (wrapped in parentheses and separated by a comma). For example, if $G_1 = \{A, B, C\}$ and $G_2 = \{1, 2, 3, 4\}$, then:

$$G_1 \times G_2 = \{(A,1), (A,2), (A,3), (A,4), (B,1), (B,2), (B,3), (B,4), (C,1), (C,2), (C,3), (C,4)\}.$$

Thus, $G_1 \times G_2$ catalogs all ways to pair a red $G_1$ member with a green $G_2$ member. The coloring is optional, but is helpful for understanding the idea here. If the red letters are men dancers and the green numbers are women dancers, then $G_1 \times G_2$ catalogs all ways in which a mixed-gender dance couple could be formed. In the last example, $G_1$ had 3 members, $G_2$ had 4 members, and $G_1 \times G_2$ had 3 × 4=12 members. More generally, the size of the product of two finite groups equals the product of their sizes.

We will denote the algebraic operation in $G_1 \times G_2$ as "•." How do you think it works? You simply do the red $G_1$ part and the green $G_2$ part separately. For example, to find (5,**H**)•(7,**V**) in the group **Z** × $\mathbf{D_4}$, the answer is (5+7,**H*****V**) = (12,$\mathbf{R_{180}}$).

For more practice, let us construct the Cayley table for the product of the groups

$\mathbf{C_3}$ = {0, 1, 2} and $\mathbf{C_2}$ = {0, 1}.

The product of these two groups has the following six members:

$\mathbf{C_3} \times \mathbf{C_2}$ = {(0,0), (0,1), (1,0), (1,1), (2,0), (2,1)}.

The Cayley table looks like this:

| • | (0,0) | (0,1) | (1,0) | (1,1) | (2,0) | (2,1) |
|---|---|---|---|---|---|---|
| (0,0) | (0,0) | (0,1) | (1,0) | (1,1) | (2,0) | (2,1) |
| (0,1) | (0,1) | (0,0) | (1,1) | (1,0) | (2,1) | (2,0) |
| (1,0) | (1,0) | (1,1) | (2,0) | (2,1) | (0,0) | (0,1) |
| (1,1) | (1,1) | (1,0) | (2,1) | (2,0) | (0,1) | (0,0) |
| (2,0) | (2,0) | (2,1) | (0,0) | (0,1) | (1,0) | (1,1) |
| (2,1) | (2,1) | (2,0) | (0,1) | (0,0) | (1,1) | (1,0) |

The yellow cell in the above table represents (1,1) • (1,0) = (2,1). Just think of this equation as a red equation (1 + 1 = 2 in $\mathbf{C_3}$)

packaged together with a green equation (1 + 0 = 1 in $\mathbf{C_2}$). Product groups may be large, but they are simple. Since their algebraic operation involves doing two unrelated things, product groups are no more complicated than the smaller groups out of which they are built.

Why are product groups important? For starters, when we study the symmetry group of a new shape, we usually ask whether it is isomorphic to any familiar group. Now that the product of two familiar groups is familiar to you, you have a larger library of familiar groups against which to compare a new symmetry group. Also, if you discover that a symmetry group is isomorphic to a product group, then this discovery often helps you better understand its underlying algebraic structure.

For example, we will explain why $\mathbf{D_2} = \{\mathbf{I}, \mathbf{H}, \mathbf{V}, \mathbf{R_{180}}\}$ is isomorphic to $\mathbf{C_2} \times \mathbf{C_2} = \{(0,0), (1,0), (0,1), (1,1)\}$. Their Cayley tables look like this:

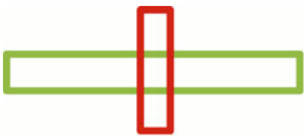
| | **I** | **H** | **V** | $\mathbf{R_{180}}$ |
|---|---|---|---|---|
| **I** | **I** | **H** | **V** | $\mathbf{R_{180}}$ |
| **H** | **H** | **I** | $\mathbf{R_{180}}$ | **V** |
| **V** | **V** | $\mathbf{R_{180}}$ | **I** | **H** |
| $\mathbf{R_{180}}$ | $\mathbf{R_{180}}$ | **V** | **H** | **I** |

| | (0,0) | (1,0) | (0,1) | (1,1) |
|---|---|---|---|---|
| (0,0) | (0,0) | (1,0) | (0,1) | (1,1) |
| (1,0) | (1,0) | (0,0) | (1,1) | (0,1) |
| (0,1) | (0,1) | (1,1) | (0,0) | (1,0) |
| (1,1) | (1,1) | (0,1) | (1,0) | (0,0) |

The isomorphism is the following one-to-one correspondence:

$$\mathbf{I} \leftrightarrow (0,0),\ \mathbf{H} \leftrightarrow (1,0),\ \mathbf{V} \leftrightarrow (0,1),\ \mathbf{R_{180}} \leftrightarrow (1,1).$$

It is simple to check that this dictionary translates each of the above Cayley tables into the other and is therefore an isomorphism.



Even better, there is a nice visual way to see this isomorphism. Think of $\mathbf{D_2}$ as the symmetry group of this red and green cross. Notice that **H** exchanges the ends of the red rectangle, **V** exchanges the ends of the green rectangle, and $\mathbf{R_{180}}$ exchanges both. In $\mathbf{C_2} \times \mathbf{C_2}$, think of the red number as recording whether the ends of the red rectangle were exchanged (0 = no, 1 = yes), and think of the green number as recording whether the ends of the green rectangle were exchanged. Thus, $\mathbf{R_{180}} \leftrightarrow (1,1)$ because the 180° rotation exchanges the ends of both rectangles. The group $\mathbf{D_2}$ appears simpler once you notice that computing in this group really just involves simultaneously keeping track of the answers to two yes/no questions. Doing two unrelated things simultaneously is what product groups are all about.

Here is a second visual illustration of the utility of product groups. The symmetry group of the B border pattern includes translations and the reflection over the horizontal center line.

… B B B B B B B B B B B B B B B B B …

My favorite symmetry of this border pattern is ($T_{-5}$,YES), which is my shorthand for "translate 5 letters to the left and flip over the horizontal center line." My second favorite symmetry is ($T_8$,NO), which is my shorthand for "translate 8 letters to the right and do not flip over the horizontal center line." Composing symmetries is easy. For example, ($T_{-5}$,YES) • ($T_8$,NO) = ($T_3$,YES). Similarly, ($T_6$,YES) • ($T_4$,YES) = ($T_{10}$,NO). Two unrelated things are happening simultaneously here: the T-subscripts are added and the YES/NOs are combined according to the table at the right.

| | NO | YES |
|---|---|---|
| NO | NO | YES |
| YES | YES | NO |

Since these two things are unrelated, we suspect that the symmetry group of the B border pattern is isomorphic to a product group. In fact, since the above table looks just like the Cayley table for $\mathbf{C_2} = \{0,1\}$ (with 0 = NO and 1 = YES), the symmetry group of the B-frieze is isomorphic to $\mathbf{Z} \times \mathbf{C_2}$. The isomorphism matches the symmetry ($T_n$,NO) with (n,0) and matches

$(T_n, \text{YES})$ with (n,1). Do you see why this is an isomorphism? For example, watch this isomorphism translate a true equation into a true equation:

$$\begin{array}{ccccc} (T_8,\text{YES}) & * & (T_7,\text{YES}) & = & (T_{15},\text{NO}) \\ \downarrow & & \downarrow & & \downarrow \\ (8,1) & \bullet & (7,1) & = & (15,0) \end{array}$$

Do you see why it will translate every true equation into a true equation? What about an equation with two NOs? Or with one YES and one NO? In summary, the symmetry group of the B border pattern is really a product group in disguise, and this discovery helps simplify your understand the symmetry group.

## Exercises

(1) Find as many subgroups of $\mathbf{D_4}$ as you can.

(2) Recall that a square can be oriented so that only the rotations in $\mathbf{D_4}$ are symmetries of the resulting image (not the flips). For each subgroup of $\mathbf{D_4}$ which you found, can you find a way to ornament/decorate the square to make that subgroup become the symmetry group of the resulting image?

(3) In the group $\mathbf{C_3} \times \mathbf{C_2}$, find a subgroup which is isomorphic to $\mathbf{C_3}$ and a subgroup which is isomorphic to $\mathbf{C_2}$.

(4) Find the subgroup generated by each member of the cyclic groups $\mathbf{C_4}$, $\mathbf{C_5}$, $\mathbf{C_6}$, …, $\mathbf{C_{11}}$. Make a conjecture about the possible orders of members of $\mathbf{C_n}$.

(✪5) In a finite group, the LAST thing in the list <A> = {A, A • A, A • A • A, …} is always $A^{-1}$. The next-to-last thing is always

$A^{-1} \bullet A^{-1}$, and so on. Verify this is true in of all your lists in problem (4). How does this help justify the claim in the chapter that “the things obtained by composing $A^{-1}$ with itself repeatedly are redundant to the things obtained by composing A with itself repeatedly”?

(6) Make a complete Cayley table for $\mathbf{C_3} \times \mathbf{C_3}$.

(7) Explain why $\mathbf{C_{12}}$ is NOT isomorphic to $\mathbf{C_2} \times \mathbf{C_6}$, even though these two groups both have order 12.

(8) In the group $\mathbf{C_7} \times \mathbf{C_{10}}$, list the members of <(0,2)> and <(1,5)>.

(9) Explain why the proper symmetries of any object in the plane form a subgroup of its symmetry group. This subgroup is called the <u>proper symmetry group</u> of the object.

# 6. Permutations

Each symmetry of a square permutes the square's vertices, and also permutes its edges. You can tell which symmetry was performed by observing how the vertices or edges were permuted. For this reason and for many more to come, permutations are a crucial key to understanding symmetry.

A "permutation" means a rearrangement. For example, there are six permutations of the letters A, B, and C; namely:

ABC, ACB, BAC, BCA, CAB, CBA

Think of A, B, and C as letters on a magnet board, and think of a permutation as a word (not necessarily a real English word) which you can spell using all of the letters.

> **DEFINITION:** When $n \geq 2$, the collection of all permutations of $n$ ordered things is denoted $\mathbf{P_n}$ and is called the $n$th permutation group.

We will soon see that $\mathbf{P_n}$ is a group, as its name suggests.

Usually the "ordered things" are letters ordered alphabetically. For example, $\mathbf{P_3}$ = {ABC, ACB, BAC, BCA, CAB, CBA} has six members. How many permutations does {A, B, C, D, E, F} have? In other words, what is the size of $\mathbf{P_6}$? To spell a

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_6,

six-letter word on the magnet board, you have six choices for the first letter, then five choices left for the second letter, then four choices left for the next letter, and so on. The number of ways in which you can make these choices in succession equals: $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$. Therefore, the size of $\mathbf{P_6}$ equals 720. The general rule is:

**THEOREM:** the size of $\mathbf{P_n}$ equals $n!$

The symbol "$n!$" means the product of all of the integers between 1 and $n$. It is pronounced "$n$ factorial". Here are the first few:

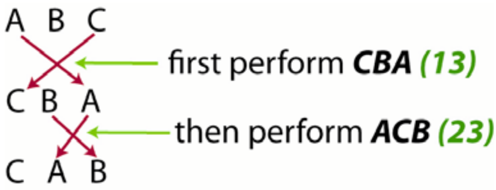
$$2! = 1 \times 2 = 2$$
$$3! = 1 \times 2 \times 3 = 6$$
$$4! = 1 \times 2 \times 3 \times 4 = 24$$
$$5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$$
$$6! = 1 \times 2 \times 3 \times 4 \times 5 \times 6 = 720$$

Factorials grow large very quickly.

## Permutation Groups

Listing and counting permutations is not nearly enough. To uncover the full power of permutations, we must learn the algebraic operation which turns $\mathbf{P_n}$ into a group. This operation is called "composition" and is denoted "*" (exactly as in symmetry groups).

To compose permutations, the key is to regard each permutation, not as a word, but as the action (exchanging/ moving/cycling of letters) that occurred to build that word on the magnet board from the (ordered) starting position. For example, in $\mathbf{P_3}$, the starting position of the magnet board is alphabetical, ABC, so this word is the identity. Each other word was obtained from this starting position by moving/exchanging/cycling some letters. For example, CBA was obtained by exchanging the 1st and 3rd letters, encoded by writing (13). ACB was obtained by exchanging the 2nd and 3rd letters, encoded by writing (23). Their composition, ACB * CBA, simply means the result of performing these exchanges in succession: first perform CBA (exchange the 1st and 3rd letters) and then perform ACB (exchange the 2nd and 3rd letters), like this:

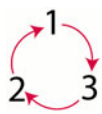



We learn that: ACB * CBA = CAB.

This equation might seem like a dreamlike intermixing of math class with spelling class. The algebra of permutations forces us to constantly alternate between two points of view: regarding a permutation as a word and as an action. To summarize:

**DEFINITION:** If $W_1$ and $W_2$ are permutations (words), then their composition, $W_1 * W_2$, is the permutation (word) obtained from the starting position by first performing the action for $W_2$ and then performing the action for $W_1$.

The shorthand symbols (13) and (23) were useful above, so let us similarly encode the remaining words in $\mathbf{P_3}$. For example, the word BCA is obtained by moving the 1st letter (A) to the 3rd position, bumping the original 3rd letter (C) to the 2nd position, and finally bumping the original 2nd letter (B) back to the 1st position. We'll encode this as (132). The red circle diagram on the right is a better code for this, but it uses too much typesetting space. The other words in $\mathbf{P_3}$ are coded as follows:




| | | |
|---|---|---|
| ABC ↔ I | BCA ↔ (132) | CAB ↔ (123) |
| BAC ↔ (12) | ACB ↔ (23) | CBA ↔ (13) |

**THEOREM: $P_n$** is a group.

This is not hard to prove. For example, every permutation has an inverse because, no matter how I scramble the letters on a magnet board, it is always possible for you to unscramble them back to the starting position. Let us practice by constructing a Cayley table for $\mathbf{P_3}$.

| * | ABC | BCA | CAB | BAC | ACB | CBA |
|---|---|---|---|---|---|---|
| **ABC** | | | | | | |
| **BCA** | | | | | | |
| **CAB** | | | | | | |
| **BAC** | | | | | | |
| **ACB** | | | | | | CAB |
| **CBA** | | | | | | |

A Cayley table for $P_3$

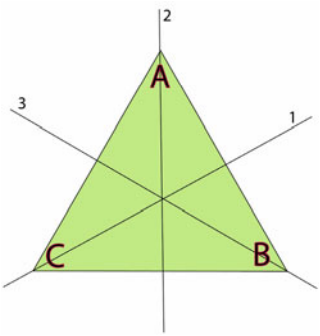
The yellow cell represents the equation ACB * CBA = CAB, which in our green "action notation" looks like (23) * (13) = (123). Fill in the rest of the table. Along the way you will have to think about both the words and the actions that the green shorthand symbols represent. When you are done, your table should look suspiciously like the Cayley table for $\mathbf{D_3}$ from Chap. 2. In fact:

**THEOREM: $P_3$** is isomorphic to $\mathbf{D_3}$.

**PROOF:** The isomorphism from $\mathbf{P_3}$ to $\mathbf{D_3}$ is:

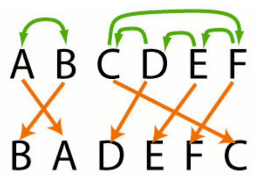
ABC ↔ **I** BCA ↔ $\mathbf{R_{120}}$ CAB ↔ $\mathbf{R_{240}}$
BAC ↔ $\mathbf{F_1}$ ACB ↔ $\mathbf{F_2}$ CBA ↔ $\mathbf{F_3}$.

This dictionary translates the Cayley table for $\mathbf{P_3}$ above into the Cayley table for $\mathbf{D_3}$ in Chap. 2. There is an important visual way to understand the isomorphism. Instead of writing three-letter words left-to-right, let us write them clockwise around the vertices of a triangular magnet board, starting at the top. For example, the identity word, ABC, is pictured here. Our isomorphism simply matches a symmetry of the triangle with the word that we see after that symmetry is performed (always reading words clockwise starting at the top). That is, we match each symmetry of the triangle with the permutation of its vertices that the symmetry induces. The permutation induced by a composition of two symmetries equals the composition of the individual permutations which these two symmetries induce. That is why it is an isomorphism.................................................................□




For $n > 3$, $\mathbf{P_n}$ is NOT isomorphic to $\mathbf{D_n}$ because $\mathbf{P_n}$ has far more members than $\mathbf{D_n}$. The issue here is that not every permutation of the vertices of an *n*-gon can be achieved by a symmetry. For example, no symmetry of a square could exchange its top two vertices without also exchanging its bottom two vertices.

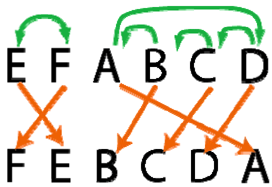
The green action symbols for $\mathbf{P_3}$ are called <u>cycle notation</u>. This is a notation system (not just in $\mathbf{P_3}$ but also in $\mathbf{P_n}$) for describing the action required to build each word from the starting position on a magnet board. For example, in $\mathbf{P_6}$, let us translate the word BADEFC into cycle notation. On a magnet board, this word is obtained from the following action:




Its cycle notation is (12)(3654), which means: exchange the 1st and the 2nd letters, and then cycle the remaining letters by sending 3rd → 6th → 5th → 4th → 3rd. Notice how the closing ")" represents an instruction to loop back to the starting "(". By the way, here is some self-explanatory jargon: (12)(3654) contains a cycle of length 2, namely (12), and a cycle of length 4, namely (3654).

You can translate any word into cycle notation by following these steps: First write "(" followed by the position of the first letter that moves. For example, if A moves, then you will write "(1". Next write the position that first letter moves to, then write the position that the letter in that position moves to, and so on. As soon as a letter moves back to the position of the first letter you considered, write ")". If any remaining letters move, then write another "(" followed by the smallest moving position which you have not yet considered, followed by the position that letter moves to and so on.

For additional practice with composing symmetries, let us find the composition BADEFC * EFABCD in **$P_6$**. For this, we use EFABCD as the starting position of the magnet board, and then perform the above-pictured action for BADEFC = (12)(3654):




We learn that in **$P_6$**, BADEFC * EFABCD = FEBCDA. Try pronouncing that!

## Even and Odd Permutations

Other than the identity, the simplest permutations are the "swaps"

**DEFINITION:** A swap means an exchange of two letters.

For example, ABEDCF is a swap in $\mathbf{P_6}$. Its cycle notation is (35) because it was obtained from the alphabetical starting position by swapping (exchanging) the 3rd and 5th letters. The cycle notation for any swap is a single cycle of length 2.

**THEOREM:** The swaps generate $\mathbf{P_n}$. In other words, every permutation in $\mathbf{P_n}$ can be expressed as a composition of swaps.

**PROOF:** To warm up, let us think about how to write the word EADCFB in $\mathbf{P_6}$ as a composition of swaps. With our magnet board in its starting position, ABCDEF, we must perform a sequence of swaps (letter exchanges) to arrive at the ending position EADCFB. We will work left-to-right, moving the E to the first position, then the A to the second, and so on:

ABCDEF → EBCDAF
EBCDAF → EACDBF
EACDBF → EADCBF
EADCBF → EADCFB.

The colored letters are being swapped. We started with positions 1 & 5 colored, then 2 & 5, then 3 & 4, and finally 5 & 6. Thus:

EADCFB = (56) * (34) * (25) * (15).

We have successfully expressed EADBFC as a composition of four swaps in $\mathbf{P_6}$. The same strategy will work to write any permutation in $\mathbf{P_n}$ as a composition of swaps.................................□

In the EADCFB example above, what if we instead work right-to-left? How many swaps are needed? What if we only use adjacent swaps (which means exchanges of adjacent letters)? What if we correctly position the vowels first and then the consonants? Try several different strategies, and count the number of swaps needed for each strategy. Some strategies require four swaps. Others might require 6 or 8 or 10 swaps.

A grossly inefficient strategy might require 76 swaps. Here is the important point. Sit down first because it is the most important thing you will ever learn. No matter what strategy you employ, you will require an *even* number of swaps. This is because EADCFB is an “even permutation.”

**DEFINITION:** A permutation that can be obtained by composing an even number of swaps is called an even permutation. A permutation that can be obtained by composing an odd number of swaps is called an odd permutation.

The importance stems from the following theorem:

**THEOREM:** A permutation cannot be both even and odd.

In other words, if you obtain your favorite word in $\mathbf{P_n}$ using an even number of swaps, then this word is even – anyone else would also require an even number of swaps to obtain it. We will not discuss the proof of this theorem, which is a bit difficult, but please do spend some time investigating it. For example, verify using several different strategies that CADEFB requires an odd numbers of swaps and is, therefore, an odd permutation.

**THEOREM:** Exactly half of the permutations in $\mathbf{P_n}$ are even. Furthermore, the even permutations form a subgroup of $\mathbf{P_n}$.

**PROOF:** To show that the even permutations form a subgroup, we must verify three things. First, the identity is an even permutation because we can obtain it using 0 swaps, and 0 is an even number. Second, the composition of two even permutations is an even permutation. For example, if the first permutation uses 8 swaps and the second uses 12 swaps, then the composition uses 8 + 12 = 20 swaps (the sum of two even numbers is even). Finally, the inverse of an even permutation is even. For example, the inverse of (56) * (34) * (25) * (15) is (15) * (25) * (34) * (56) – the same swaps done in the reverse order. If your friend performs a sequence of swaps, you can return to the starting position by undoing your friend's swaps in reverse order. This verifies that the even permutations form a subgroup of $\mathbf{P_n}$. In the exercises, you will be asked to prove that exactly half of them are even……..□

**DEFINITION:** The subgroup of all even permutations in $\mathbf{P_n}$ is denoted $\mathbf{A_n}$ and is called the *n*th alternating group.

The size of $\mathbf{A_n}$ equals half the size of $\mathbf{P_n}$. The sizes of the first few permutation and alternating groups are:

| | |
|---|---|
| Size of $\mathbf{P_2}$ = 2 | Size of $\mathbf{A_2}$ = 1 |
| Size of $\mathbf{P_3}$ = 6 | Size of $\mathbf{A_3}$ = 3 |
| Size of $\mathbf{P_4}$ = 24 | Size of $\mathbf{A_4}$ = 12 |
| Size of $\mathbf{P_5}$ = 120 | Size of $\mathbf{A_5}$ = 60. |

The groups $\mathbf{P_4}$, $\mathbf{A_4}$, and $\mathbf{A_5}$ will play starring roles in the next chapter, so remember their sizes, and keep an eye out for them.

We end this chapter with a final useful observation:

**THEOREM:** A cycle of length *m* can be obtained by composing *m* – 1 swaps. Thus, the cycle is even if *m* is odd, and viceversa.

For example, the word BCDEFA in $\mathbf{P_6}$ is expressed in cycle notation as (165432). This is a cycle of length 6. Using a magnet board, check that BCDEFA = (56) * (45) * (34) * (23) * (12), so it is obtained from five swaps. These five swaps "bubble" the first letter A to the end of the word.

This theorem is useful for quickly deciding whether a permutation is even or odd. Using cycle notation is often faster than using a magnet board. For example, we previously decided that BADEFC = (12)(3654). The first cycle has length 2, so it is odd. The second cycle has length 4, so it is odd. BADEFC is, therefore, the composition of an odd cycle with an odd cycle, which is even.

## Exercises

(1) If you have not yet done so, fill in the Cayley table for $\mathbf{P_3}$ in the chapter.

(2) Express each of these words in $\mathbf{P_6}$ in cycle notation, and decide whether each is even or odd: CBAFDE, BCDAFE, FABCDE, EDAFCB.

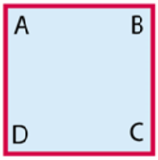
(3) Find CBAFDE * BCDAFE and FABCDE * EDAFCB in $\mathbf{P_6}$.

(4) Find the five-letter word corresponding with each of these cycle notation expressions in $\mathbf{P_5}$, and decide whether each is even or odd: (13)(245), (13524), (13)(45).

(5) What is the *order* of a cycle of length m?

(6) List the six members of $\mathbf{P_3}$, and circle the even ones. Via the isomorphism between $\mathbf{P_3}$ and $\mathbf{D_3}$, which symmetries of the triangle are matched with even permutations?

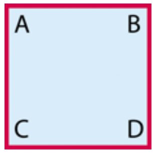
(7) List all 24 permutations of the letters {A,B,C,D}. Write each permutation in cycle notation. Circle the even ones.




(8) Label the vertices of a square as shown. In your list of the 24 permutations of {A,B,C,D}, decide which permutation is induced by each of the eight symmetries of the square. That is, after performing the symmetry, which word do you see (always

reading words clockwise around the square starting at the top-left vertex)? Verify that the symmetries which induce even permutations are exactly {**I**, **H**, **V**, $\mathbf{R_{180}}$}.

(9) In the previous exercise, instead of reading words clockwise around the square, what if we read in the "page of a book" order, so that the identity position is the one illustrated here. Verify that the symmetries which induce even permutations are {**I**, **H**, **V**, $\mathbf{R_{180}}$}, as before.




*The punch line: It makes sense to ask whether a symmetry induces an even or odd permutation of the vertices; the answer does not depend on how the vertices were ordered.*

(10) Which of the symmetries of a regular pentagon induce even permutations of its vertices? What about a regular hexagon? Conjecture a pattern: which symmetries of a regular *n*-gon induce even permutations of its vertices? *HINT: Consider separately the cases where n is even and where n is odd.*

(11) Explain why the adjacent swaps generate $\mathbf{P_n}$. In other words, every permutation in $\mathbf{P_n}$ can be expressed as a composition of adjacent swaps.

(12) Express the identity permutation in $\mathbf{P_7}$ in several different ways as a composition of swaps. Check that you always have an even number of swaps.

(13) Prove that exactly half of the permutations in $\mathbf{P_n}$ are even. *HINT: use the "ancient scroll" idea in the proof of the All-Or-Half Theorem from Chap. 2.*

(14) On the final exam for their symmetry course, students are asked to match five objects with their five symmetry groups. What do you think is the fairest method for grading such a matching problem? For example, try regarding each student's answer as a

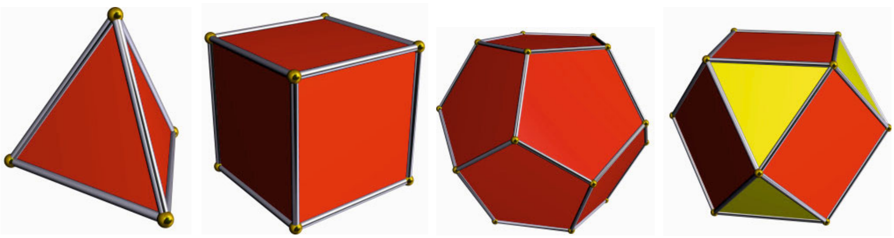
permutation in $\mathbf{P_5}$, and scoring based on how “close” it is to the identity permutation (the correct answer). How should “close” be measured?

(15) If H is a subgroup of $\mathbf{P_n}$, prove that either all or half of the members of H are even.

(16) Is $\mathbf{P_2}$ isomorphic to $\mathbf{C_2}$?

# 7. Symmetries of Solid Objects

So far, we have only studied the symmetries of flat two-dimensional objects in the plane. In this chapter, we discuss the symmetries of solid three-dimensional objects in space. Look at the following solid objects, and rank them from the most symmetric to the least symmetric.



Images from Robert Webb's Great Stella software, *http://www.software3d.com/Stella.html*

Each of these objects is built from red and yellow faces, silver edges, and gold vertices. To rank them, we must decide how our previous methods for studying symmetry apply to solid objects.

## Rigid Motions of Space

A symmetry of a solid object is defined in the same way that a symmetry of a flat object was defined:

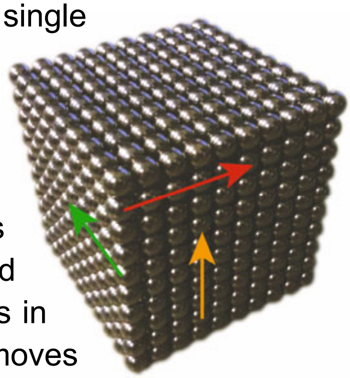
> **DEFINITION:** A symmetry of a solid object in space means a rigid motion of space which leaves the object apparently unchanged.

But what is a "rigid motion of space"? Intuitively, it is a moving/ repositioning of *all* of space that does not compress, expand, or otherwise distort distances. Imagine that space is completely filled with transparent ice, with an object entombed somewhere

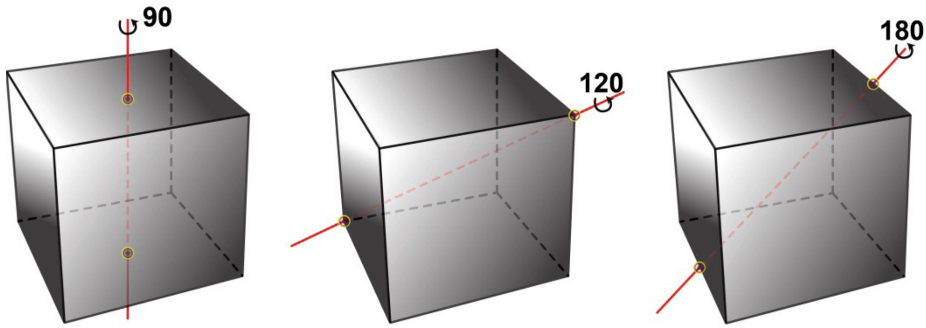
K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_7,

within. A rigid motion moves the entire infinite expanse of ice, and it is called a symmetry of the object if it leaves the object apparently unchanged. How can an infinite expanse of ice be moved, and where does the mover stand? We do not have a good answer; the ice image is imperfect, but it is the best we can offer until we provide a more precise definition of a rigid motion in a later chapter. For now, here are some examples.

**TRANSLATIONS:** A translation moves every point of space the direction and distance specified by a single arrow. Imagine that the ball pattern illustrated here is continued infinitely in all directions, so that the balls fill up all of space. Each of the three arrows represents a translation symmetry of this infinite ball pattern. For example, the red arrow translates each ball 5 ball positions in the right-back direction. Since each ball moves exactly on top of another ball, an observer would not notice that the translation occurred – that is what makes it a symmetry.



**ROTATIONS:** In space, we rotate about an axis (which means a line), not about a point. Imagine children rotating about a maypole. The axes of some rotation symmetries of a cube are shown below. Try to do each of these rotations to a cardboard cube to convince yourself that each one really is a symmetry. Entombing your cardboard cube in an infinite expanse of ice is optional; it is fine here to think of each rotation as something done just to the cube rather than to all of the surrounding space.




3 axes of rotation symmetries of the cube

The order of an axis means the order of the smallest non-identity rotation symmetry about that axis. For example, the axes pictured above have orders 4, 3, and 2, respectively.

**REFLECTIONS:** A reflection across a plane is visualized by thinking of the plane as a mirror, and moving each point of space (each speck of ice) to the position of its mirror image on the opposite side of the mirror. For example, the reflection across the green plane exchanges the ice above and below this plane; it is NOT a symmetry of the human figure because, after this reflection occurs, he would appear upside down. Neither is the reflection across the blue plane, which would leave him facing backwards. Only the reflection across the orange plane is a symmetry of the human figure.

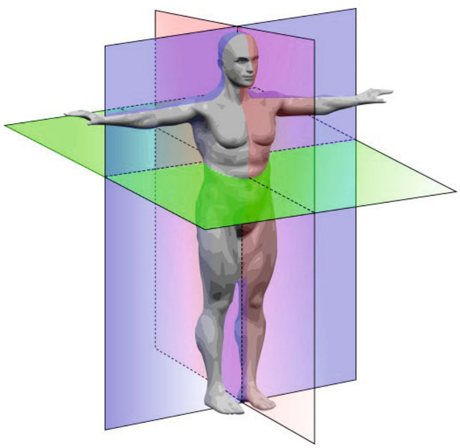


Image by YassineMrabet, Wikipedia.org

It is easy to visualize a translation or a rotation being physically done to a solid object (and to the surrounding ice, if you wish). However, reflections are different. A reflection cannot be physically done to a solid object. There is nothing you can physically do to the human figure that will exchange his right and left halves. The reflected image of his right hand looks like his left hand, but there is nothing you can physically do to a solid right hand to turn it into a solid left hand. Why should we care about reflections if we cannot do them? Even though reflections cannot be physically performed, they still help explain why solid objects appear to be symmetric. The man has bilateral symmetry, which means that his only symmetry (other than the identity) is a single reflection. That is why his left half looks the same as (or at least like the mirror image of) his right half. What other living things have approximately bilateral symmetry?

A rigid motion is called proper, if it transforms a solid right hand into a solid right hand and thus *can* be physically done to solid objects. Rotations and translations are proper. A rigid motion is called improper, if it transforms a solid right hand into a solid left hand, and thus cannot be physically done to solid objects. A reflection is improper, and so is a reflection followed by a rotation or translation.

We will eventually define "rigid motion of space" more precisely, but to get by for now, we offer this:

**CLASSIFICATION OF RIGID MOTIONS OF SPACE:** Any rigid motion of space can be obtained by composing rotations, reflections, and translations (no more than one of each kind is needed).

In other words, there are no rigid motions of space other than the types we just considered (and compositions thereof).

An object in space is called bounded if it is entirely contained in some cube. Otherwise, it is called unbounded. The meaning would remain unaltered if the word "cube" was replaced by "sphere" or "pyramid" or many other possibilities; if an object can be fit inside one of these shapes, then it can be fit into all of them. The previously mentioned infinite ball pattern was unbounded; we used it to illustrate translation symmetries. *Bounded* object never have translation symmetries. In fact:

**THE CENTER POINT THEOREM:** Any bounded solid object in space has a "center point" such that:

**(1)** Every proper symmetry is a rotation about an axis through this center point and

**(2)** Every improper symmetry is a reflection across a plane through this center point, possibly composed with a rotation about an axis through this center point.

Part (2) is necessarily lengthier than its analog for bounded flat objects, because for bounded solid objects, a reflection composed with a rotation is NOT necessarily equal to a reflection.

The symmetries of any solid object form a group (under the operation of composition). As before, this group is called the symmetry group (or the full symmetry group) of the object. The proper symmetries form a subgroup, which is called the proper symmetry group of the object.

Our main goal in this chapter is to understand all of the possible ways in which bounded solid objects can be symmetric. But what exactly does this mean? When should we consider two bounded solid objects to be "symmetric in the same way"? Should this mean that they have isomorphic symmetry groups? Or should it mean that the have isomorphic *proper* symmetry groups? Or should it mean that they are rigidly equivalent or perhaps properly rigidly equivalent, defined as follows:

**DEFINITION:** Two solid objects are called rigidly equivalent if there exists a rigid motion of space which, when applied to the first object, repositions it so that afterwards, the two objects have exactly the same symmetries.

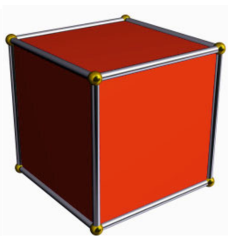
Two solid objects are called properly rigidly equivalent if there exists a rigid motion of space which, when applied to the first object, repositions it so that afterwards, the two objects have exactly the same *proper* symmetries.

These are all reasonable ways to capture the idea that a pair of objects is "symmetric in the same way." Some imply others. For example, if two objects are rigidly equivalent, then their full symmetry groups are isomorphic (just as before). Similarly, if two objects are *properly* rigidly equivalent, then their *proper* symmetry groups are isomorphic.

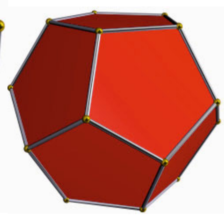
Why are *proper* symmetries getting so much attention? Because understanding an object's proper symmetries is often easier than understanding all of its symmetries. You will see. The Center Point Theorem tells us that the proper symmetries of a solid bounded object are all rotations.

Our next goal is to understand the full symmetry group (or at least the proper symmetry group) of a tetrahedron, cube, and dodecahedron pictured below.

Tetrahedron



Cube



Dodecahedron

Take time now to build these three shapes. Cardboard works well. To keep track of the required building materials, we will denote by "F" the number of faces and by "S" the number of sides that each face has. A tetrahedron has F = 4 faces that are identical equilateral triangles (S = 3). A cube has F = 6 faces that are identical squares (S = 4). A dodecahedron has F = 12 faces that are identical pentagons (S = 5). Cut out the faces and glue or tape them together. Another approach is to find cut-out templates for each shape by doing a web image search for something like "dodecahedron template."

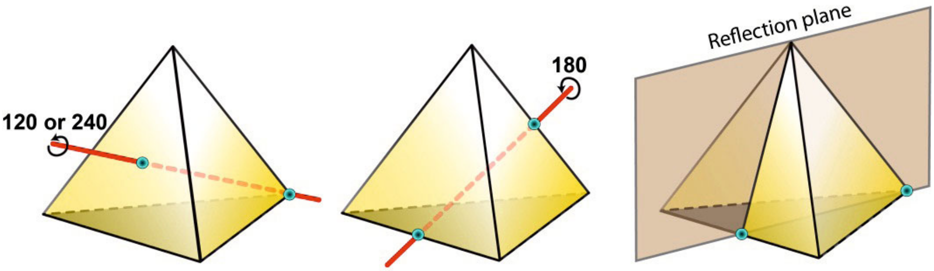
Set one of these three cardboard shapes on a table. To perform a proper symmetry, you might pick it up, rotate it, and then set it back down on the table in its original footprint. In how many ways can you do this? You really only have two choices to make: which face goes down and how this bottom face is rotated. The number of ways to make these two choices in succession equals F × S. We, therefore, expect that the proper symmetry groups of these three shapes have the following sizes:

| | F | S | Number of proper symmetries |
|---|---|---|---|
| Tetrahedron | 4 | 3 | 12 |
| Cube | 6 | 4 | 24 |
| Dodecahedron | 12 | 5 | 60 |

Counting symmetries is a good start, but it is not enough. We will now look in more depth at the symmetry group (or at least the proper symmetry group) of each of these three shapes.

## The Symmetry Group a Tetrahedron

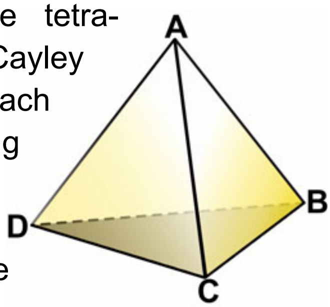
Illustrated below are three types of symmetries of a tetrahedron:




Pictured first is an axis of order 3, which connects a vertex to the midpoint of its opposite face. There are four such axes (one for each vertex), with rotations by 120 and 240° about each, yielding a total of 8 non-identity rotations about such axes.

Pictured second is an axis of order 2, which connects the midpoints of a pair of opposite edges. There are three such axes (one for each pair of edges), with a 180° rotation about each, yielding a total of 3 non-identity rotations about such axes. We thus have 8 + 3 + 1 = 12 total rotation symmetries (counting the identity), as expected. These are all of the tetrahedron's proper symmetries.

Pictured third is a reflection plane. There are six such planes, one for each edge. These are all of the tetrahedron's reflection planes; however, they might NOT be all of the tetrahedron's improper symmetries. The Center Point Theorem does NOT say that every improper symmetry is a reflection.

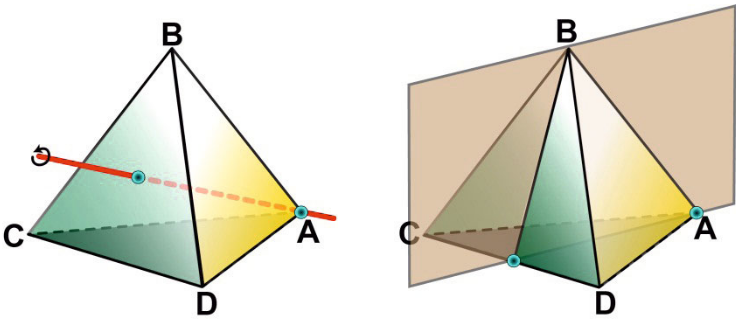



We wish to fully understand the tetrahedron's symmetry group. To build a Cayley table, we must first choose a name for each symmetry. Here is an illuminating naming system. Let us label the four vertices A, B, C and D as illustrated – think of this illustration as the identity position, which we call ABCD. Name each other symmetry with the four-letter word that you see after that symmetry is performed. Always read words in the order indicated by the illustration (top, bottom-right, bottom-front, and bottom-left). Thus, we name symmetries according to the permutations of the tetrahedron's vertices which they induce. This system will inevitably lead you to stumble upon the following theorem:

**THEOREM:** A tetrahedron's symmetry group is isomorphic to $\mathbf{P_4}$, and its proper symmetry group is isomorphic to $\mathbf{A_4}$.

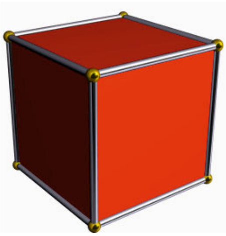
**PROOF:** The previously described naming strategy associates each symmetry with a permutation in $\mathbf{P_4}$ (a four letter word). No two different symmetries could induce the same permutation; that is, if you know how the vertices were permuted, then you know what symmetry was performed.

Further, for any permutation of the vertices, we can find a symmetry that achieves that permutation. How? First, the top vertex (A) can easily be rotated into the correct position. In both illustrations below, it has just been rotated to the bottom-right position. Next, the rotation and reflection pictured below can be combined to obtain all six possible permutations of the remaining three vertices (B, C, and D around the green face).




In summary, there is a one-to-one correspondence between the symmetries of the tetrahedron and the permutations of its four vertices. This one-to-one correspondence provides the desired isomorphism. The permutation induced by the composition of two symmetries equals the composition of the permutations which these two symmetries induce. That is why it is an isomorphism. In the exercises, you will verify that the *proper* symmetries correspond to the *even* permutations, which is why the proper symmetry group is isomorphic to $\mathbf{A_4}$.............................□

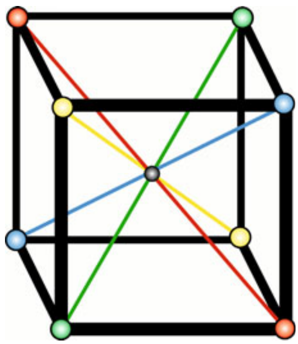
## The Proper Symmetry Group a Cube



Next, let us study the *proper* symmetry group of a cube (we will worry about the *full* symmetry group later). The cube has 24 proper symmetries. The only permutation group or alternating group with this size is $\mathbf{P_4}$, which might lead you to guess the following:

**THEOREM:** The proper symmetry group of a cube is isomorphic to $\mathbf{P_4}$.

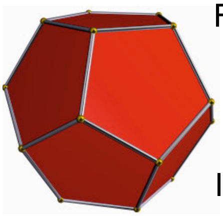
**PROOF:** A cube has 6 faces, 8 vertices, and 12 edges. To mimic the idea of the previous proof, we must find something of which the cube has 4. The solution is pictured here; it has 4 diagonals (colored green, blue, red, and yellow). The four diagonals are permuted by the proper symmetries of the cube. Check that any permutation of these four diagonals can be achieved by a proper symmetry. Also check that no two different proper symmetries ever induce the same permutation of the colors. You will need a physical cube to convince yourself of these claims. If your cube is solid, you cannot run colored diagonals through it, but you could instead just color its vertices. In summary, there is a one-to-one correspondence between



the proper symmetries of the cube and the permutations of its four diagonals. This correspondence is the desired isomorphism……………………………………………………………□

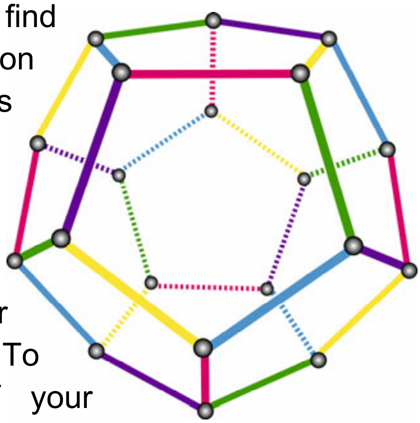
Here is an alternative phrasing of the key idea from the previous proof. The largest possible stick (line segment) that is able to fit inside of a cube can actually fit in exactly four different ways, and these four ways are permuted by the symmetries of the cube.

## The Proper Symmetry Group a Dodecahedron



Finally, let us turn our attention to the dodecahedron. The dodecahedron has 60 proper symmetries. The only permutation group or alternating group with this size is $\mathbf{A_5}$, which might lead you to conjecture that:

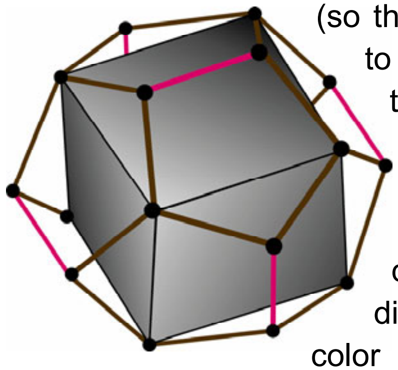
**THEOREM:** The proper symmetry group of the dodecahedron is isomorphic to $\mathbf{A_5}$.



**IDEA OF PROOF**: We need to find something of which the dodecahedron has five, which get permuted by its proper symmetries. Here is the answer: it is possible to color all of its edges using five colors (blue, green, pink, purple, and yellow) in such a way that each proper symmetry permutes these colors. To verify this, color the edges of your cardboard model as illustrated. If you color correctly, then you can verify the following: (1) Each proper symmetry permutes the colors; for example, if a proper symmetry sends one pink edge to a yellow edge, then it sends all pink edges to yellow edges. (2) Each *proper* symmetry permutes the colors by an *even* permutation. (3) No two different proper symmetries ever permute the colors in the same way. (4) Every *even* permutation of the colors can be achieved by a *proper* symmetry. Use your cardboard model to make sure you understand and believe each of these claims.......................................................................□

Here are explicit instructions for coloring your dodecahedron model to make the previous proof work. You need a total of six pink edges. Color some first edge pink, and the remaining five pink edges should be the edges that are either parallel or perpendicular to the first. Repeat with the other colors.

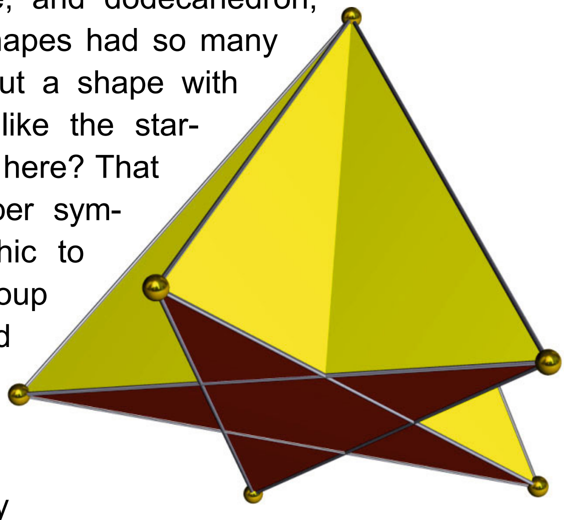
Since the six edges of a single color are chosen to be mutually parallel or perpendicular, each color determines a "frame." Each such color frame instructs you how to position a cube that you might wish to inscribe inside of the dodecahedron (so that every edge of the cube is parallel to an edge of the dodecahedron with that color). The illustration on the left shows a cube positioned by the pink frame. In fact, a largest possible cube that is able to fit inside of a dodecahedron can actually fit in five different ways (determined by the five color frames), and these five ways are permuted by the symmetries of the dodecahedron.



This draws a nice analogy between the cube's story and the dodecahedron's story. Fitting largest possible sticks into a cube is analogous to fitting largest possible cubes into a dodecahedron.

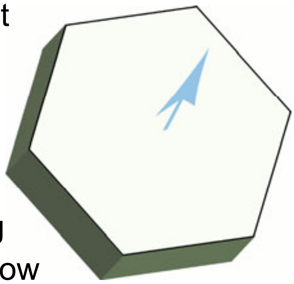
## Solid Objects Which Are "Essentially Two-Dimensional"

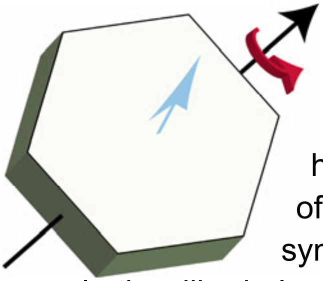
It was hard work figuring out the proper symmetry groups of the tetrahedron, cube, and dodecahedron, largely because these shapes had so many rotation axes. What about a shape with only one rotation axis, like the star-based pyramid illustrated here? That is much easier. Its proper symmetry group is isomorphic to the proper symmetry group of the star-shaped shadow that it would cast on the ground, namely $\mathbf{C_5}$. We will call this pyramid "essentially two-dimensional" because its proper symmetries are no more complicated than the symmetries of its two-dimensional shadow (cast when light is aimed down its only rotation axis).



Images from Robert Webb's Great Stella software
http://www.software3d.com/Stella.html

The thick hexagon pictured on the right is also "essentially two-dimensional." At first glance, it might appear that the central blue axis (which has order 6) is its only rotation axis, but look again! In fact, it has six other axes, each of order 2. We are talking about the side axes, like the one drawn below in black.





Like all rotations, the 180° side axis rotation is a *proper* symmetry, which might seem confusing because it reminds you of an *improper* symmetry of the flat hexagon. Thus, the *proper* symmetry group of the thick hexagon is isomorphic to the *full* symmetry group of its flat hexagon shadow, namely the dihedral group $\mathbf{D_6}$. In fact, the word "dihedral" comes from the Greek "dihedron" which means a solid with two faces, like the thick hexagon. Whenever a flat two-dimensional object is thickened a bit, its *proper* symmetry group becomes isomorphic to the flat object's *full* symmetry group. This is because the third dimension of space can be used to rotate the object upside down.

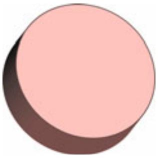
**DEFINITION:** A solid three-dimensional object is called essentially two-dimensional if it has no more than one rotation axis of order >2.

Thus, an essentially two-dimensional solid object only has one rotation axis (if you do not count order 2 axes). The solid object's proper symmetries are no more interesting than the symmetries of the two-dimensional shadow cast when light is shown down this main symmetry axis.

More specifically, the star-based pyramid and the thick hexagon above exemplify the only two ways that an essentially two-dimensional solid object can be related its shadow: (1) There are no side axes, and the proper symmetry group of the solid is isomorphic to the *proper* symmetry group of its shadow or (2) There are side axes, and the proper symmetry group of the solid is isomorphic to the *full* symmetry group of the shadow. In either case, Da Vinci's Theorem implies:

**THEOREM:** The proper symmetry group of any essentially two-dimensional solid object is either infinite or isomorphic to a dihedral or cyclic group.

The hockey puck and the cone shown here are examples of essentially two-dimensional solid objects with infinitely many symmetries. Each has a circular shadow. The proper symmetry group of the hockey puck is isomorphic to the full symmetry group of this circular shadow. The proper symmetry group of the cone is isomorphic to the proper symmetry group of this circular shadow.

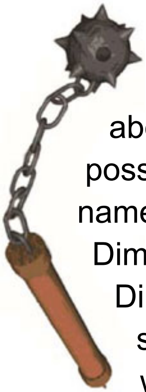


## The Classification Theorem for Bounded Objects

In the last several sections, each new example was more complicated than the previous. The tetrahedron had 12 proper symmetries, the cube had 24, and the dodecahedron had 60. Can we continue building more and more complicated solid objects with more and more symmetries? We did so with flat objects when we considered 2-gons, 3-gons, 4-gons, 5-gons…this list goes on indefinitely, with each shape having more symmetries than the previous one. But quite surprisingly, we cannot build solid objects which are any more complicated than the objects that we have already considered! The three-dimensional analog of Da Vinci's theorem is much more restrictive that you might have expected. Here it is:

**THE CLASSIFICATION THEOREM:** Every bounded solid object is either essentially two-dimensional or is properly rigidly equivalent to a tetrahedron, cube, dodecahedron, or sphere.

If the object is essentially two-dimensional, then its proper symmetry group must be infinite or isomorphic to a cyclic or dihedral group. If the object is genuinely three-dimensional, it is amazing that there are only four more possibilities for its proper symmetry group! Thus, the group version of the classification theorem says this: The proper symmetry group of any bounded solid object is infinite or isomorphic to a dihedral or cyclic group or $\mathbf{A}_4$, $\mathbf{P}_4$, or $\mathbf{A}_5$.

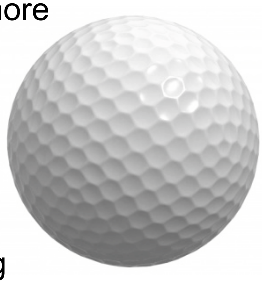


It is remarkable that there are so few possibilities for the proper symmetry groups of solid bounded objects! To appreciate what this means, let us think about people who might wish that there were more possibilities. For example, imagine a medieval blacksmith named Robin who is commissioned by a knight named Sir Dim to build a spike ball weapon with 12 sharp spikes. Sir Dim insists that the 12 spikes be "spread out with no bald spots," to best smite his enemies. That is, he wants the weapon to look the same to an enemy facing one spike as it does to an enemy facing any other spike. He is really requesting here that the spike arrangement be symmetric – ignoring the chain, there should be a proper symmetry moving any spike to any other spike. Since a dodecahedron has 12 faces, Robin is in luck – he can arrange the 12 spikes at the midpoints of the faces of a dodecahedron.

The next year, Sir Dim demands a new weapon built with 20 spikes. Robin is again in luck, because the dodecahedron has 20 vertices, so he can arrange the spikes at the vertices of a dodecahedron. Robin could also manage 30 spikes, arranged at the midpoints of the 30 edges of the dodecahedron, or even 60 spikes (think about how). But the next year, Sir Dim requests a weapon with 100 spikes – the largest number he knows, and Robin is out of luck. It is impossible to build a perfectly symmetric spike ball with more than 60 spikes because such a weapon would have more proper symmetries than a dodecahedron, contradicting the classification theorem. The only way it could be done is with an essentially two-dimensional arrangement of spikes around the equator of the ball, but this arrangement would have large bald spots, and is not what Sir Dim had in mind.

This is not an engineering restriction or an issue of manufacturing imperfections – it is a purely mathematical limitation. There is a limit to the number of points which can be symmetrically arranged around a sphere! Making the sphere larger or the points smaller does not help. It just cannot be done.

There are many modern versions of this tale. For example, a golf ball engineer is unable to distribute more than 60 dimples symmetrically around a golf ball, and must, therefore, settle for an *approximately* symmetric arrangement of the 250–450 dimples found on most balls today. Unfortunately, this means that the aerodynamics could in principle change depending on how the ball is set on the tee. On the other hand, a lack of dimple symmetry is not all bad. Certain intentionally asymmetric dimple patterns, like those used on Polara brand balls, can cause a ball to fly straighter. The USGA banned the use in tournament play of asymmetric Polara balls, which lead Polara to file a lawsuit against the USGA. The USGA was in an interesting legal and mathematical situation here; it was trying to regulate that balls must be as symmetric as possible, even while the classification theorem says that no ball could ever be perfectly symmetric.



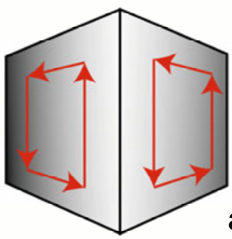
Architects are also limited by the classification theorem. For example, the arrangement of triangles around a geodesic dome structure like the Epcot center is *approximately* symmetric at best. Similarly, the arrangement of more than 60 seeds around a goats beard or dandelion could not be perfectly symmetric. In biology, proteins have been discovered representing all possible types of proper symmetry groups allowed by the classification theorem: cyclic, dihedral, $\mathbf{A_4}$, $\mathbf{P_4}$, and $\mathbf{A_5}$. It may not be obvious how to account for the symmetry of a particular protein, but the serious lack of options is certainly part of the answer.

Epcot Center photo by Ylebru, Wikipedia.org

Goats beard photo by Ken Tapp

## Chirality

A solid object is called <u>chiral</u> if all of its symmetries are proper. For example, the decorated cube pictured on the left is chiral; after a reflection, it would look different because the red cycle on each face would turn clockwise instead of counterclockwise, so a reflection could not be a symmetry. Hold your book up to a mirror to see what we mean. Imagine that the cube and its reflection are both solid objects (not flat images on the page and on the mirror surface). Do you see how they differ?

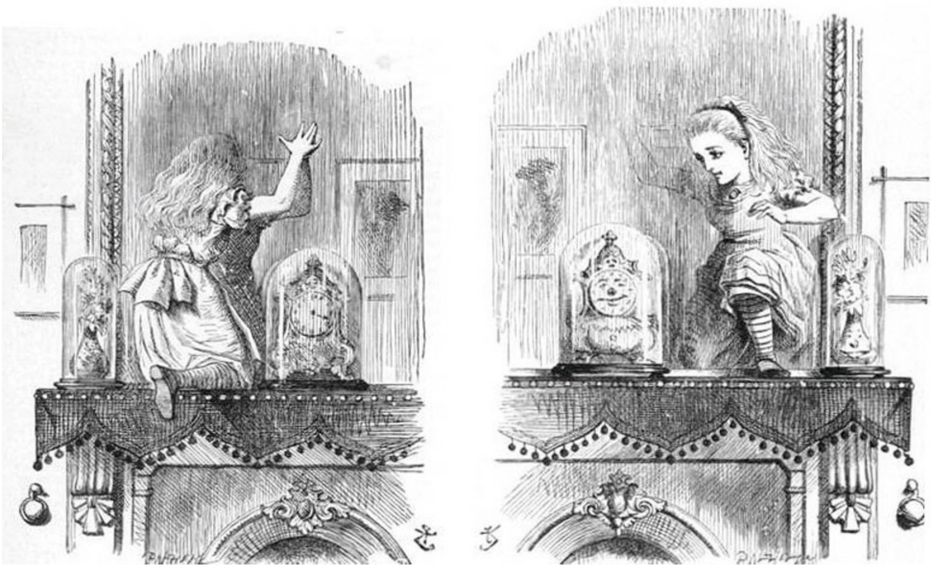


You could create a chiral tetrahedron or dodecahedron in the analogous way. A more artistic variation is shown on the right.

Bulatov's metal sculpture is approximately a chiral dodecahedron; it has the same proper symmetry group as a dodecahedron, but it has no improper symmetries. Hold it up to a mirror to check this.

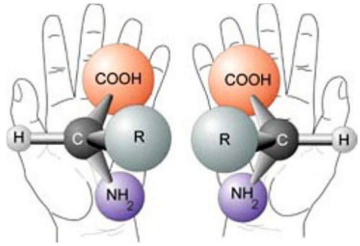
*Dodecahedron XIV* metal sculpture by Vladimir Bulatov (bulatov.org)

"Looking-glass" is an old English term for a mirror. In the opening scene of Lewis Carroll's novel, *Through the Looking Glass and What Alice Found There*, Alice imagines that the reflected image of her room in the Looking glass is part of a real other world. She asks her cat, "How would you like to live in Looking-glass House, Kitty? I wonder if they'd give you milk in there? Perhaps Looking-glass milk isn't good to drink." This comment is far more interesting that it might at first appear.



Alice goes through the looking glass

All chiral objects act and look differently in Looking-glass world. For example, the writing on books is backwards. Also, Looking-glass screws are “counter screws,” so Looking-glass carpenters learn the rule “righty loosey, lefty tighty,” which does not even rhyme. This makes it difficult to graduate from Looking-glass carpenter school.




More serious differences lurk at the microscopic level. Many molecules are chiral, which means that their reflections in Looking-glass world will look and act differently. In the illustration here, the reflected image of the left-handed amino acid is the right-handed amino acid. In fact, the word “chiral” derives from the Greek word for “hand.” Naturally occurring organic chiral molecules are almost always found only in the left- or right-hand version, not both. The right-handed version of a molecule may interact differently with mechanisms in living cells that evolved to interact with the left-hand version. For example, the molecule *carvone* is responsible for the smell of caraway seeds, but the mirror-image version of carvone smells like spearmint!

*Thalidomide* is a deadly example of the difference between the left- and right-hand versions of molecules. It was prescribed in the late 1950s to control morning sickness, but it turned out to cause birth defects. Before being recalled in 1961, thalidomide was responsible for birth defects in tens of thousands of infants worldwide. This tragedy may have been caused by the manner in which thalidomide was produced. The synthetic manufacturing of chiral molecules yields equal amounts of the left- and right-hand versions of the molecule. It is thought that one version of thalidomide controls morning sickness, while the other causes birth defects and that the two were not properly separated in the manufacturing process. Evidently, it is not safe for Kitty to drink Looking-glass milk.

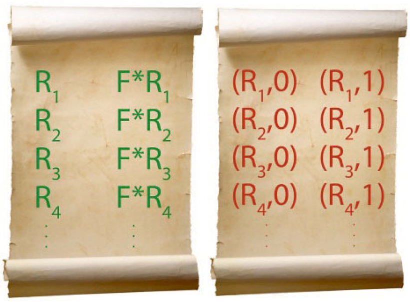
## Proper Versus Full Symmetry Groups

We have so far only classified the *proper* symmetry groups of bounded solid objects. What about the *full* symmetry groups? The All-Or-Half Theorem and its proof are valid for solid objects (check this). That is, either all or half of a solid object's symmetries are proper. In certain situations, it is easy to explain how the proper and improper symmetries fit together into a single Cayley table.

**THEOREM:** If an object (two-dimensional or three-dimensional) has a single improper symmetry of order 2 that commutes with every one of its proper symmetries, then its full symmetry group is isomorphic to the product of its proper symmetry group and $\mathbf{C_2}$.

A symmetry, **A**, is said to commute with another symmetry, **B**, if the order in which the pair are performed does not matter; that is, **A** * **B** = **B** * **A**.

**PROOF IDEA:** Suppose there is an improper symmetry of order 2 that commutes with all proper symmetries and call it F. Name the proper symmetries as $R_1$, $R_2$, $R_3$, and so on. Recall from the proof of the All-Or-Half Theorem that the green scroll pictured here is a listing of all of the object's symmetries (proper on the left and improper on the right). The red scroll is a listing of all things in the product of the object's proper symmetry group and $\mathbf{C_2}$. The two scrolls are aligned so as to suggest a one-to-one correspondence. We match the left of the green scroll with the left of the red scroll, so that $R_n \leftrightarrow (R_n,0)$. We match the right of the green scroll with the

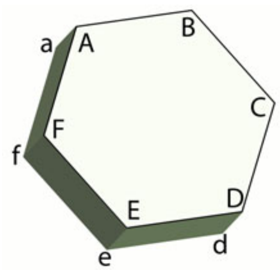


right of the red scroll, so that $F * R_n \leftrightarrow (R_n, 1)$. Think about why this pairing is an isomorphism. That is, why does it translate every true green equation into a true red equation? To answer this, separately consider green equations that combine a pair of green symmetries from the left, the right, or one of each………..□

**NON-EXAMPLE:** In $\mathbf{D_4}$, there is no flip that commutes with all of the rotations, or even just with $\mathbf{R_{90}}$ (check this using the Cayley table in Chap. 2). Thus, the square does not satisfy the hypothesis of the theorem, nor does it satisfy the conclusion, since $\mathbf{D_4}$ is NOT isomorphic to $\mathbf{C_4} \times \mathbf{C_2}$.

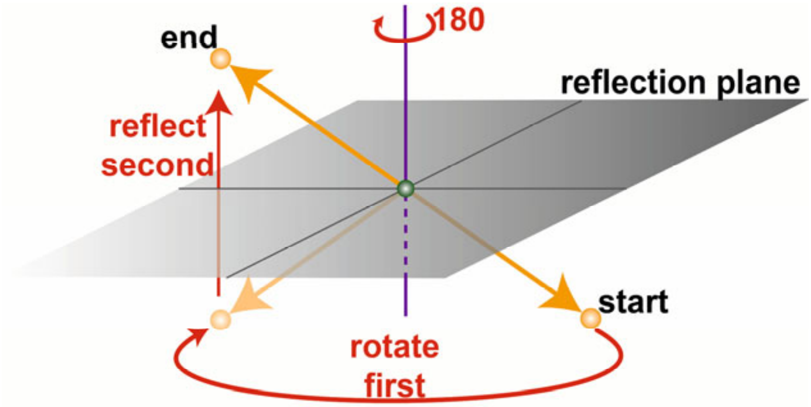
**EXAMPLE:** At the end of Chap. 5, we discovered that the symmetry group of the B border pattern is isomorphic to $\mathbf{Z} \times \mathbf{C_2}$. Since the horizontal flip commutes with each translation, the above theorem leads us to this same conclusion. Compare our reasoning in Chap. 5 to the logic of the above proof.

**EXAMPLE:** We previously discovered that the proper symmetry group of a thick hexagon is isomorphic to $\mathbf{D_6}$. We now claim that its full symmetry group is isomorphic to $\mathbf{D_6} \times \mathbf{C_2}$. To justify this claim, we must find an improper symmetry which commutes with all 12 proper symmetries. The answer is the reflection across the plane that is parallel to the top and bottom hexagons, and half way between them. With the 12 vertices labeled as shown, this reflection exchanges lower case with upper case letters, which might help you visualize why it commutes with the 12 rotations.




As for the cube and dodecahedron, each one of these has a special improper symmetry called <u>central inversion</u> that commutes with all of its proper symmetries. Therefore, the full symmetry group of the cube is isomorphic to $\mathbf{P_4} \times \mathbf{C_2}$, and the full symmetry group of the dodecahedron is isomorphic to $\mathbf{A_5} \times \mathbf{C_2}$.

Central inversion is similar to a reflection, but instead of reflecting across a plane, it reflects across a point, namely the center point of the object. What does it mean to "reflect across a point"? It means to do two things in succession: First rotate 180° about an axis through that point and then reflect across the plane through that point that is perpendicular to this axis. In the diagram below, central inversion across the green point moves the each position (like the yellow one) to its "antipode" across this green point.




Central inversion moves the yellow start point to its antipode across the green center-point

An object is called centrally symmetric if central inversion (across its center point) is a symmetry of the object. When we later learn about matrices, we will verify that central inversion always commutes with every other symmetry of a centrally symmetric object. For now, can you visualize why this might be true? Thus, the full symmetry group of any centrally symmetric object is isomorphic to the product of its proper symmetry group with $\mathbf{C_2}$. Can you visualize why cubes and dodecahedrons are centrally symmetric, but tetrahedrons are not?

You might conjecture that any bounded solid object (that is not essentially two-dimensional) is *rigidly equivalent* to one these seven objects: tetrahedron, cube, dodecahedron, chiral tetrahedron, chiral cube, chiral dodecahedron, sphere. This is a good guess, but to make it true, an 8th object must be added to the list; namely, a volleyball. The volleyball illustrated on the right has the same proper symmetry group as a tetrahedron, but it differs from the tetrahedron in that it is centrally symmetric, so its full symmetry group is isomorphic to $\mathbf{A_4} \times \mathbf{C_2}$. In summary, any bounded solid object (that is not essentially two-dimensional) is *rigidly equivalent* to a sphere or to one of these objects:

| | Proper sym. group. | Full sym. group. |
|---|---|---|
| Tetrahedron | $\mathbf{A_4}$ (12) | $\mathbf{P_4}$ (24) |
| Chiral tetrahedron | | $\mathbf{A_4}$ (12) |
| Volleyball | | $\mathbf{A_4} \times \mathbf{C_2}$ (24) |
| Cube | $\mathbf{P_4}$ (24) | $\mathbf{P_4} \times \mathbf{C_2}$ (48) |
| Chiral cube | | $\mathbf{P_4}$ (24) |
| Dodecahedron | $\mathbf{A_5}$ (60) | $\mathbf{A_5} \times \mathbf{C_2}$ (120) |
| Chiral dodecahedron | | $\mathbf{A_5}$ (60) |

Notice that we have not yet described the possible full symmetry groups of essentially two-dimensional solid objects. This topic will be discussed in the exercises. It will complete our picture of all possible ways in which bounded solid objects can be symmetric.

## Exercises

(1) Identify the proper and full symmetry group of several of Vladimir Bulatov's sculptures displayed at *http://bulatov.org*. Which are chiral?

(2) Verify that all 12 *proper* symmetries of the tetrahedron permute the vertices by *even* permutations; thus, the proper symmetry group of a tetrahedron is isomorphic to $\mathbf{A_4}$.

(3) Is every improper symmetry of the tetrahedron equal to a reflection?

(4) Describe the full symmetry groups of the star-based pyramid, the hockey puck and the cone (pictured in this chapter). Which of these objects are centrally symmetric?

(5) Fill in the following blanks:

The tetrahedron has ___ axes of order 3 and ___ axes of order 2.

The cube has ___ axes of order 4, ___ axes of order 3, and ___ axes of order 2.

The dodecahedron has __ axes of order 5, ___ axes of order 3, and ___ axes of order 2.

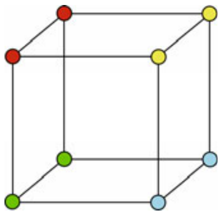
Using nothing but this information, how could you count the proper symmetries of each object?

(6) We saw that the symmetry group of the tetrahedron is $\mathbf{P_4}$, thought of as the group of all permutations of its four vertices. Could you alternatively think of this as the group of all permutations of its four faces? In other words, can we replace vertices with faces in the proof?

(7) Explain what is wrong with the following reasoning: "The six faces of the cube get permuted by each symmetry of the cube; therefore, the symmetry group of the cube is isomorphic to $\mathbf{P_6}$."

(8) Explain what is wrong with the following reasoning: "The four colored diagonals of the cube get permuted by each symmetry of the cube; therefore, the full symmetry group of the cube is isomorphic to $\mathbf{P_4}$."

(9) Explain what is wrong with the following reasoning: "The four colors (painted on the vertices in the illustration on the right) get permuted by each proper symmetry of the cube; therefore, the proper symmetry group of the cube is isomorphic to $\mathbf{P_4}$." How could you recolor the vertices to make this reasoning correct?



(10) Determine the proper and full symmetry group of a box whose length, width, and height are all different. What if exactly two of these measurements are the same?

(11) If a solid object has an odd number of total symmetries, what is the strongest conclusion you can make about its symmetry group?

(12) Is Looking-glass milk safe for Looking-glass Kitty to drink?

(13) The faces of a soccer ball are pentagons and hexagons. Obtain a soccer ball and determine its full and proper symmetry group. *Hint: What kinds of rotation axes does it have? Compare its answer to that of the tetrahedron, cube, and dodecahedron.*

(14) Is a thick hexagon centrally symmetric? What about a thick pentagon? What about a thick *n*-gon?

(15) identify the proper and full symmetry groups of the single, double, and triple inner tubes pictured below.

(16) The seven bracelets below are obtained by wrapping the seven border patterns around cylindrical bands. Each bracelet is essentially two-dimensional.

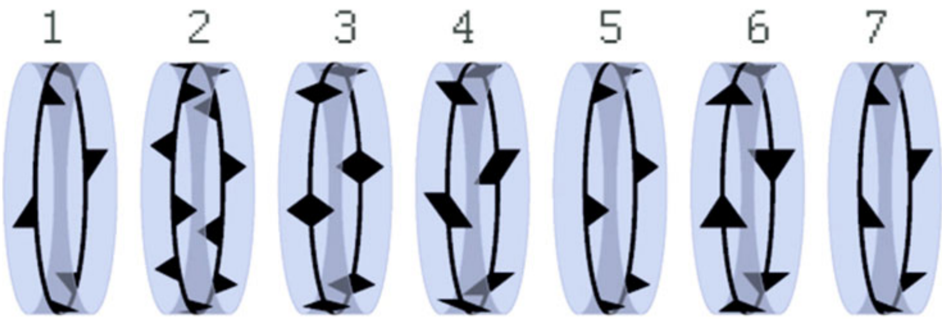



Image by AndrewKepert, Wikipedia.org

(a) Which of these bracelets are chiral?
(b) Which of these bracelets have rotation axes of order 2?
(c) The proper symmetry group of each bracelet is isomorphic to either a cyclic group or a dihedral group. Which are cyclic and which are dihedral?
(d) Identify an improper symmetry of the 6th bracelet that commutes with all of its proper symmetries.

(e) The first bracelet has exactly three proper symmetries and three improper symmetries. Describe all of them. Notice that none of the improper symmetries are reflections. What familiar group is this bracelet's symmetry group isomorphic to?

(f) Find a pair of bracelets whose full symmetry groups are isomorphic.

*COMMENT: Each bracelet was obtained from a border patterns by wrapping some number (usually 6) of iterations of the border pattern's basic design element around a band. By using different numbers of iterations, you obtain not just seven bracelets, but seven families of bracelets. These families of bracelets represent all possible ways in which essentially two-dimensional bounded objects can be symmetric. More precisely, any essentially two-dimensional bounded object with a finite symmetry group is rigidly equivalent to one of the bracelets in one of the families.*

# 8. The Five Platonic Solids

The regular polygons played the staring rolls in our study of symmetry among two-dimensional bounded objects. Though not as symmetric as a circle, they did pretty well considering they were built out of straight lines. Each corner looked the same as each other corner, and each edge looked the same as each other edge. The three-dimensional analog of a regular polygon is called a "regular polyhedron" or a "Platonic solid":

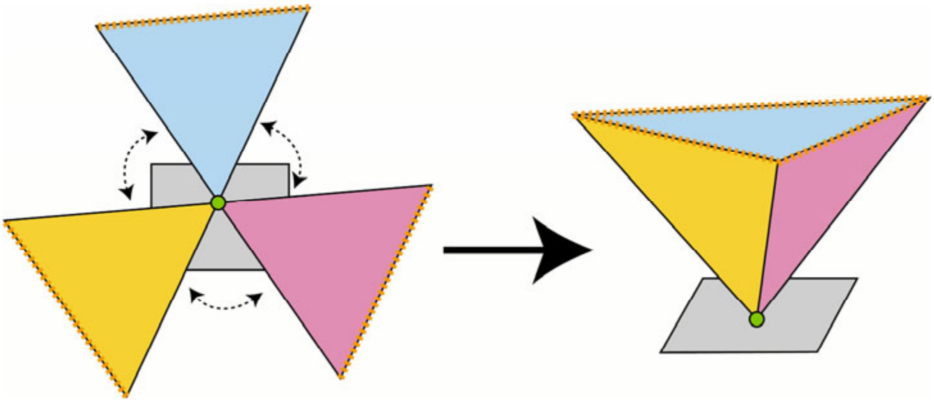
**DEFINITION:** A regular polyhedron (also called a Platonic Solid) means a bounded three-dimensional object whose faces are all identical regular polygons assembled such that each vertex has the same number of faces meeting at it.

We have already encountered three Platonic solids: the tetrahedron, cube and dodecahedron. How many others are there? Since there is an infinite list of regular polygons (each next one looking more like a circle), you might expect that there is also an infinite list of Platonic solids (each next one looking more like a sphere). Or perhaps you expect the opposite – that the tetrahedron, cube and dodecahedron are the only Platonic solids, since these were the only objects referenced in the classification theorem of the previous chapter. Or perhaps you suspect that there are exactly five Platonic solids because the chapter title ruined the surprise.

Whatever your guess, we will now attempt to systematically find and classify the Platonic solids. You will need some supplies. Obtain scissors and cardboard, and cut out a small pile of identical copies of each of these regular polygons: 3-gons (triangles), 4-gons (squares) and 5-gons (pentagons). Start by choosing a pile, which we will call "S" (for Sides). For example, S = 3 means you are building your Platonic solid entirely out of

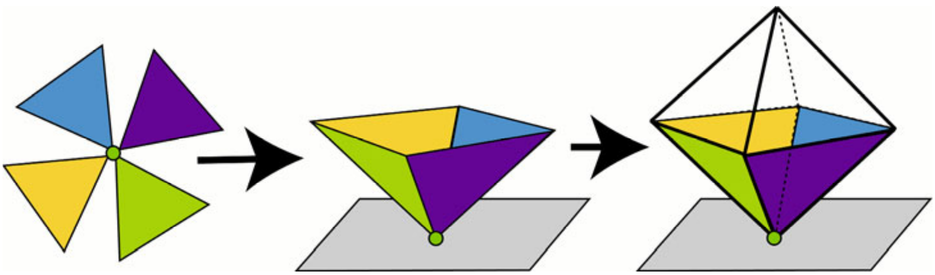
K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_8, 

triangles, S = 4 means squares, S = 5 means pentagons, and so on. Next decide how many copies of this shape will meet at each vertex. We will call this choice "C" (for Copies). Assemble C copies of an S-gon together to form your first vertex, and tape up the sides. Then do the same for each new vertex, and see whether you end up with a valid Platonic solid. For example, the choice S = 3 and C = 3 yields a tetrahedron:

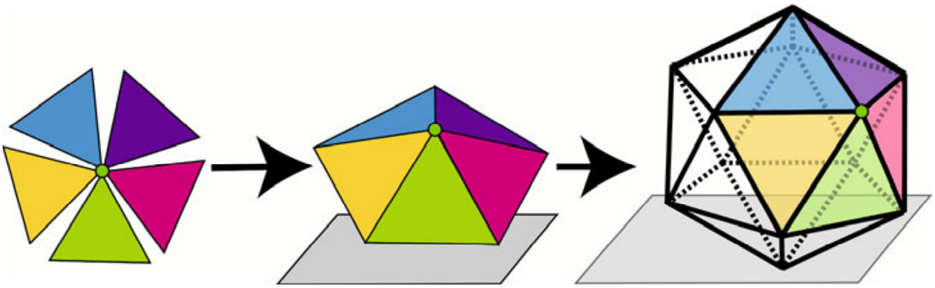


**S=3 and C=3 (triangles meeting 3 to a vertex) yields a tetrahedron.**

The tetrahedron is not the only Platonic solid that you can build with triangles. The choices C = 4 and C = 5 yield new Platonic solids called the octahedron and the icosahedron.

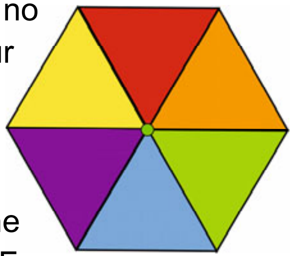


**S=3 and C=4 (triangles meeting 4 to a vertex) yields an octahedron.**

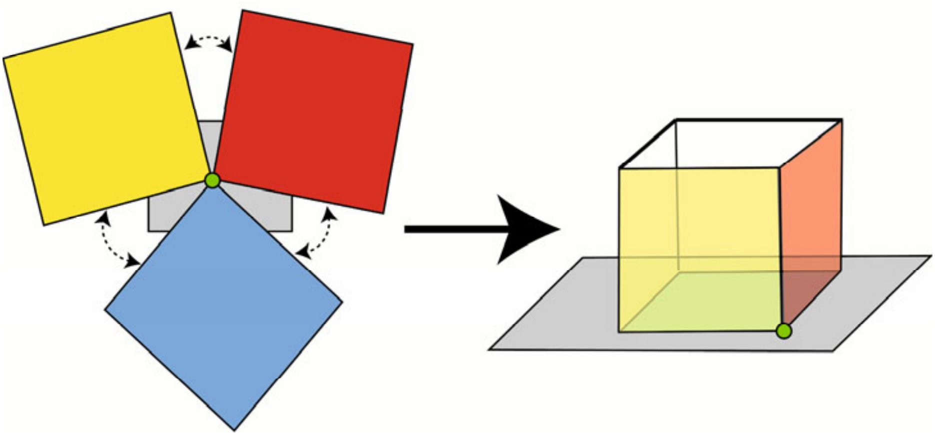


**S=3 and C=5 (triangles meeting 5 to a vertex) yields an icosahedron.**

What else can be built with triangles? The choice S = 3 and C = 6 does not seem to work because six triangles meeting at a vertex leaves no extra room. There are no gaps between the six triangles, so your construction will never bend upwards to become a bounded solid; rather, no matter how many triangles you add, you will end up with a flat triangular tiling of a portion of the plane. What about S = 3 and C > 6? For example, try taping together seven triangles meeting at a vertex. This is actually possible to do, as long as you allow some of the tape lines to extend up above the plane and others to extend down below the plane. Try to convince yourself that this construction could never be completed to form a valid (bounded) Platonic solid. We will use the term overcrowding to refer to the phenomenon where more regular polygons meet at a vertex than can fit together in the plane. We conjecture that overcrowding could never lead to a valid Platonic solid, so we will not waste time trying any more overcrowded possibilities for S and C.



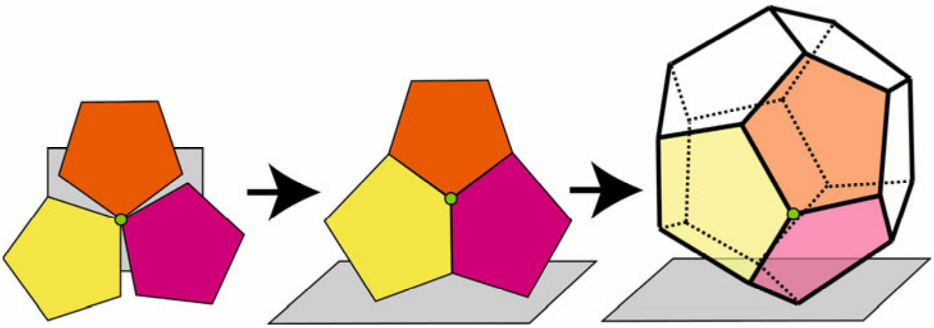
What can be built out of squares? The cube is one familiar possibility, illustrated below:



S=4 and C=3 (squares meeting 3 to a vertex) yields a cube.

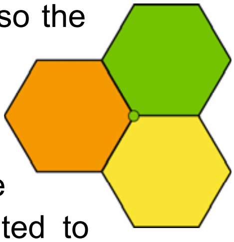
S = 4 and C = 4 does not work because four squares meeting at a vertex leaves no extra room, and would, therefore, yield a square-tiling of a portion of the plane. S = 4 and C > 4 does not work due to overcrowding.

What can be built from pentagons (S = 5)? The only possibility is C = 3, since C > 3 would be overcrowded. Thus, only the dodecahedron can be built from pentagons:

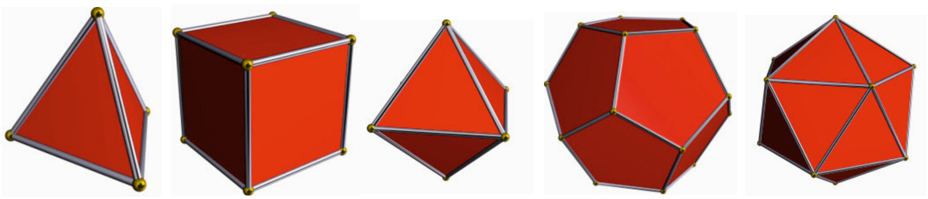


S=5 and C=3 (pentagons meeting 3 to a vertex) yields a dodecahedron.

Are there any possibilities with S > 5? Exactly three hexagons fit together in the plane with no gaps, so the choice S = 6 and C = 3 would lead to a hexagonal tiling of the plane, not a bounded Platonic solids. Any attempt to build a Platonic solid with S > 6 would fail because of overcrowding. We have arrived at an important theorem, usually attributed to Plato:



**PLATO'S THEOREM:** There are exactly five Platonic solids: the tetrahedron, cube, octahedron, dodecahedron, and icosahedron.



Tetrahedron Cube Octahedron Dodecahedron Icosahedron

Images from Robert Webb's Great Stella software, *http://www.software3d.com/Stella.html*

The above tape-and-cardboard discussion provides very strong evidence that this theorem is true, but we must acknowledge that more work is required to achieve a completely airtight proof of this theorem. In particular, it is still necessary to prove more rigorously that overcrowding never yields a valid Platonic solid. Also, it is necessary to prove that the tetrahedron, cube, octahedron, dodecahedron, and icosahedron are all perfect (not just approximate) Platonic solids. For example, what if it turned out that a dodecahedron could only be assembled using slightly irregular pentagons, whose angles vary a degree or two from being all the same? Our cardboard and tape was too sloppy to detect such a minor issue. We will not discuss these remaining details. Mathematicians have taken care of these issues, so we can move forward with assurance that the theorem is true – there are exactly five Platonic solids!

## Counting Their Parts

If you have not already done so, take time now to build the remaining Platonic solids out of cardboard. You need to build them to count their parts. Use your cardboard models to verify that the number of vertices (V), edges (E) and faces (F) for each Platonic solid is as follows:

| | V | E | F | S | C |
|---|---|---|---|---|---|
| Tetrahedron | 4 | 6 | 4 | 3 | 3 |
| Cube | 8 | 12 | 6 | 4 | 3 |
| Octahedron | 6 | 12 | 8 | 3 | 4 |
| Dodecahedron | 20 | 30 | 12 | 5 | 3 |
| Icosahedron | 12 | 30 | 20 | 3 | 5 |

As before, "S" means the number of sides that each face has, and "C" means the number of faces meeting at each vertex. Examine your Platonic solid models, and notice that "C" also equals the number of edges emanating from each vertex. That is another way to say the same thing.

In case we miss-counted the edges, let us check our work by deriving E from the other numbers in the above table. There are actually two useful methods for deriving E:

(1) What is wrong with this logic: *"Since the cube has six faces, each with four edges, the cube must have 6 × 4=24 total edges"*? Why did this computation mistakenly double the correct answer that the cube has 12 total edges? Each edge was mistakenly double-counted because each edge belongs to two faces. The correct formula is this: the number of faces times the number of edges per face equals *twice* the number of edges. In symbols: $F \times S = 2 \times E$.

(2) What is wrong with this logic: *"Since the cube has eight vertices, each with three edges emanating from it, the cube must have 8 × 3=24 total edges"*? As before, each edge was mistakenly double-counted, this time because each edge emanates from two vertices. The correct formula is: the number of vertices times the number of edges emanating from each vertex equals *twice* the number of edges. In symbols: $V \times C = 2 \times E$.

Check that the two blue-boxed formulas above are true in each row of the Platonic solid table.

## Duality

What other patterns do you notice in the Platonic solid table? The color coding is a hint. The cube and octahedron rows are colored purple. How are these rows similar to each other? The dodecahedron and icosahedron rows are colored green. How are these rows similar to each other? What patterns can you identify? Do these numerical patterns indicate a coincidence or a fundamental geometric relationship between the color-coded pairs of Platonic solids? Hoping for the latter might lead you to discover the concept of "duality."

**DEFINITION:** Duality is a procedure for starting with one Platonic solid and constructing another Platonic solid. The new solid is built from the old solid via these steps:

**1.** The vertices of the new solid are the centers of the faces of the old solid.

**2.** Two vertices of the new solid are connected via a new edge exactly when the faces of the old solid (on which these vertices are centered) share an old edge.

If the old solid is a cube, let us imagine what the new solid will look like. To help you visualize the process, find a room which is shaped roughly like a cube. Paint an “X” on the center of its roof, its floor, and each of its four walls. These X-marks are the vertices of the new solid. Use kite string for the edges of the new solid. Run a piece of kite string from the ceiling-X to each wall-X (because the ceiling shares an edge with each wall), and similarly from the floor-X to each wall-X. Run a piece of kite string between the X-marks on each pair of walls that share an edge (front-right, front-left, back-right, and back-left). Do NOT run a piece of kite string between X-marks on pairs of walls that do not share an edge (ceiling-floor, right-left, front-back). As depicted on the right, the new solid is an octahedron!

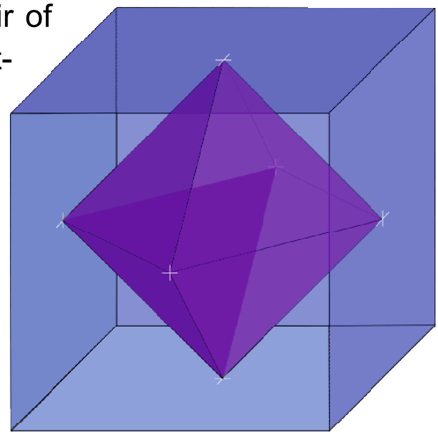
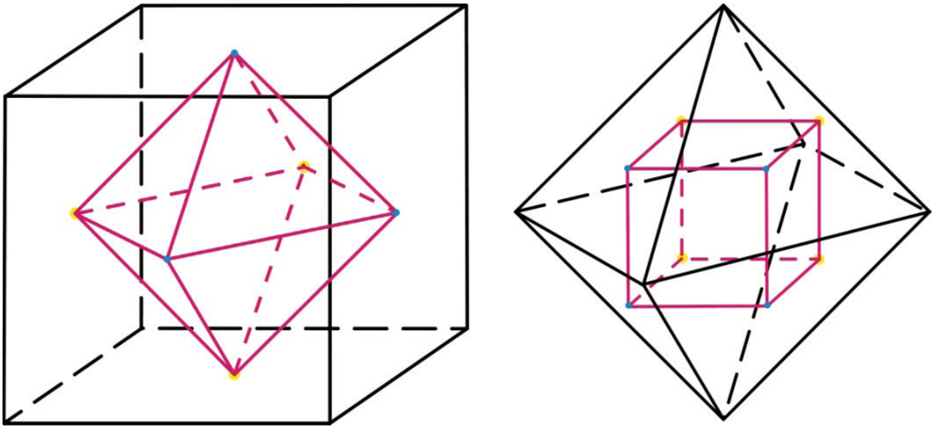


Image by User:4C, Wikipedia.org

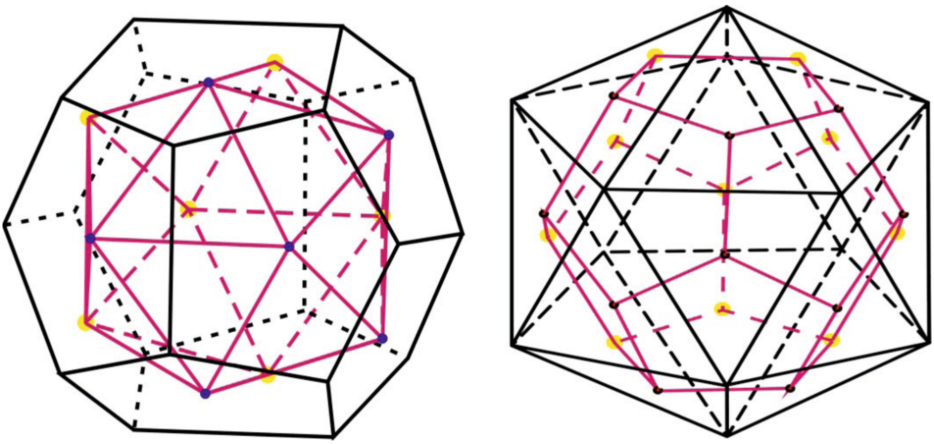
This illustration helps us account for all of the relationships that we previously discovered between the cube and octahedron rows of the table. For example, the number of vertices of the new solid (the octa-hedron) equals the number of faces of the old solid (the cube) because the new solid’s vertices are the centers of the old solid’s faces. Further, the new and old solid have the same number of edges, because the duality procedure required us to construct one (kite string) edge of the new solid for each (ceiling–wall, floor–wall, or wall–wall boundary) edge of the old solid. Finally, the number of faces of the new solid equals the number of vertices of the old solid. If eight spiders built their webs at the eight vertices of the old solid (the corners of the room), then each spider would have a different face of the new solid so spend it is day staring at.

The duality procedure undoes itself. If the old solid is the octahedron, then the new solid is the cube! We say that the cube and octahedron form a dual pair or that each is dual to the other. Imagine yourself in an octahedron-shaped room, painting X-marks and running kite string. Can you see the cube taking shape? It is illustrated below.



Cube-octahedron duality. Images by Peter Steinberg, *commons.wikimedia.org*

If you start with a dodecahedron, the duality procedure yields an icosahedron, and vise versa; these two Platonic solids form another dual pair.



Dodecahedron-Icosahedron duality. Images by Peter Steinberg, *commons.wikimedia.org*

If you start with a tetrahedron, then the duality procedure yields a tetrahedron. The tetrahedron is therefore called <u>self-dual</u>. Try to draw a picture of this.

In each of the above duality illustrations, it is easy to see that the old solid and the new solid have the same proper and improper symmetries. Every symmetry of the old solid is also a symmetry of the new solid inscribed inside it, and vice versa. In particular, the octahedron has the same (proper and full) symmetry group as the cube, while the icosahedron has the same (proper and full) symmetry group as the dodecahedron. With no extra work required, we have identified the proper and full symmetry groups of our two new Platonic solids!

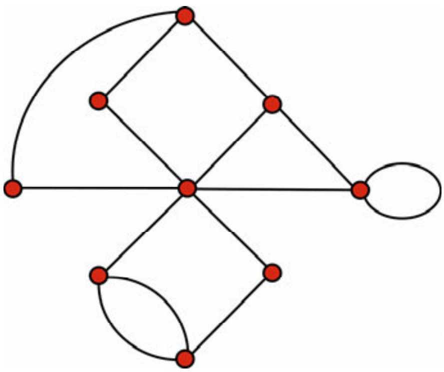
## Euler's Formula

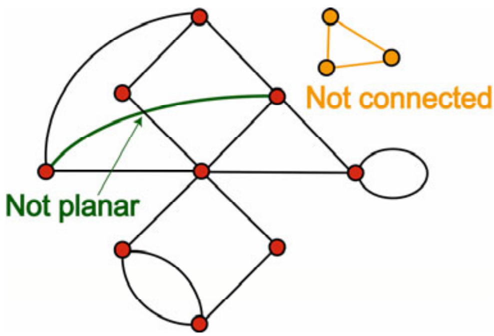
Here is a reprinting of our Platonic solid table, with two new columns added at the end to celebrate our new understanding of the proper and full symmetry groups of dual pairs, and with a mysterious new white column added in the middle.

| | V | E | F | V+F–E | S | C | Prop | Full |
|---|---|---|---|---|---|---|---|---|
| Tetrahedron | 4 | 6 | 4 | 2 | 3 | 3 | $\mathbf{A_4}$ | $\mathbf{P_4}$ |
| Cube | 8 | 12 | 6 | 2 | 4 | 3 | $\mathbf{P_4}$ | $\mathbf{P_4 \times C_2}$ |
| Octahedron | 6 | | 8 | 2 | 3 | 4 | | |
| Dodecahedron | 20 | 30 | 12 | 2 | 5 | 3 | $\mathbf{A_5}$ | $\mathbf{A_5 \times C_2}$ |
| Icosahedron | 12 | | 20 | 2 | 3 | 5 | | |

As the white column shows, each of the five Platonic solids has the mysterious property that its vertex count plus its face count minus its edge count equals 2. In symbols: $V + F - E = 2$. Why always 2? Is this a coincidence, or an indication of some underlying geometric principle waiting to be discovered? Hoping for the latter, let us hunt for the most general setting in which this formula is valid.

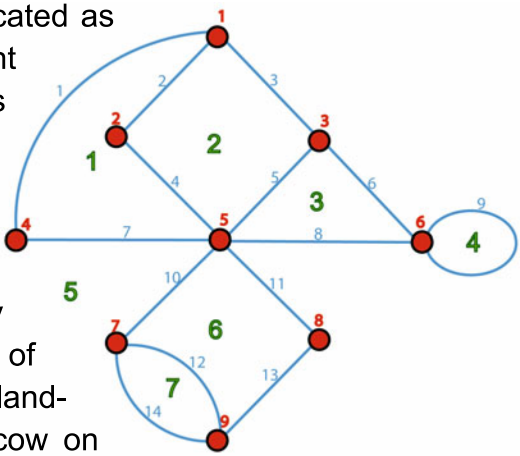
First, this formula appears in the study of “connected planar graphs.” What are those? A graph means a finite collection of vertices (dots in the plane) together with a finite collection of edges (straight or curved paths beginning and ending at the vertices). Think of the vertices as towns and the edges as roads between the towns. A graph is called connected if it is possible to travel between any pair of vertices (towns) along the edges (roads). A graph is called planar if the edges only meet each other at vertices (the roads may meet each other at towns but do not otherwise cross each other). In the illustration below, the left graph is a connected and planar graph. The right illustration show how to change it into a non-planar graph (by adding a new edge that crosses another edge at a non-vertex location) and how to change it into a non-connected graph (by adding a new cluster of edges and vertices with no bridge to the original cluster).


This graph is connected and planar



This graph is NOT

Take out a pen and paper, and draw your own connected planar graph. Make it as complicated as you like. When you finish, count the number of vertices (V), edges (E), and faces (F) that your graph has. For example, our count for the pictured graph came out V = 9, E = 14, F = 7. A face means one of the grassy pastures into which the network of towns and roads divides the land-scape. If a farmer placed one cow on each face, then the cows could not share each other's grass, because the type of cow we are talking about here is afraid to cross roads. One of the faces is always unbounded (number 5 in our graph), so one lucky cow will have an unlimited supply of grass to munch. If you count carefully, your numbers will satisfy Euler's magic formula V + F – E = 2. In words, the number of combined faces and vertices equals two more than the number of edges. This formula is true for our graph (9 + 7 – 14 = 2), and it is true for your graph also, no matter how complicated you made it. This is guaranteed by:




**EULER'S FORMULA FOR THE PLANE:** For any connected planar graph, V + F – E = 2.

**PROOF**: The theorem seems magical, but its proof is simple. Imagine putting a piece of tracing paper over the connected planar graph that you drew. The idea is to retrace your graph, one edge at a time, in such a way that V + F – E starts equal to 2 and remains equal to 2 at each step. Start by tracing any one vertex. At this point, V = 1, F = 1 and E = 0, so V + F – E = 2 for this starting graph on your tracing paper. Next trace any edge that starts at that vertex, and also trace the vertex at which this edge ends. Continue tracing one step at a time. At each step, you must trace a new edge that begins at a previously traced vertex, and also trace the vertex at which this edge ends (unless the ending vertex was previously traced). Since your graph was connected, you will be able to completely trace it in this "one edge at a time" order. The expression V + F – E started equal to 2, and we claim that it will continue to equal 2 after each new edge is traced. Why? Each time you trace a new edge, there are two possibilities:

(1) If the new edge terminates at a previously untraced vertex, then you trace it, so at this step you just added one new edge and one new vertex and no new faces.
(2) If the new edge terminates in a previously traced vertex, then this new edge will divide one large face into two smaller faces, so at this step you just added one new edge and one new face and no new vertices.

So either you increased E and V by one each, or you increased E and F by one each. Neither of these changes effects the expression V + F – E. If this expression equaled 2 before you made the change, then it equals 2 afterwards as well. As you trace the graph, the expression V + F – E starts at 2 and never changes, so after the last edge is traced, it still equals 2............□

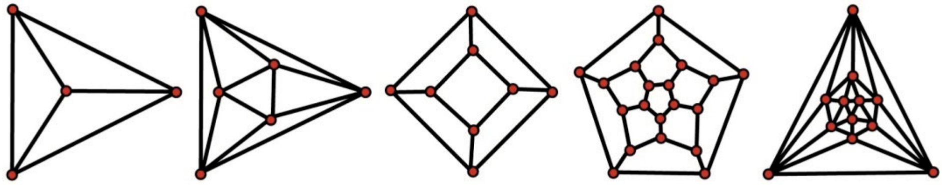
Are there other settings in which Euler's formula is true? Let us try balloons. Blow up a balloon and draw a graph on it with a permanent marker. As before, your graph must be connected, and its edges may only meet each other at its vertices. Now count V, F, and E. Does the number of combined faces and vertices equal 2 more than the number of edges? If you counted carefully, then it must!

**EULER'S FORMULA FOR THE SPHERE:** For any connected graph embedded on a sphere, V + F – E = 2.

“Embedded on a sphere” means that the graph is drawn on the surface of a sphere with its edges only meeting each other at its vertices. You might visualize it as a network of towns and roads on the surface of a planet. We do not need a new proof here; just check that the previous proof of Euler’s Theorem works equally well on a sphere.

How is Euler’s formula for Platonic solids related to Euler’s formula for the plane and the sphere? Try this. Instead of cardboard triangles, squares, and pentagons, use rubber ones to build your Platonic solids. For example, build an icosahedron out of 20 rubber triangles, carefully glued together so that the seams are airtight. Blowing air into your rubber icosahedron will balloon it into a spherical shape, and its seam lines will form a connected graph embedded on the sphere, whose vertices, edges, and faces correspond to those of the original flat-faced icosahedron. Thus, Euler’s theorem is valid for the five Platonic solids because it is valid the balloon graphs that they determine on the sphere.

Alternately, if you hold the hollow frame of a Platonic solid over a piece of paper, shine a light from above, and trace the resulting edge shadows on the paper, your tracing will be a connected graph. Choose a good angle to shine the light from so that your resulting graph is planar. After you straighten their edges and pretty them up a bit, your tracings might look like this:



Edge-shadow graphs of the 5 Platonic solids

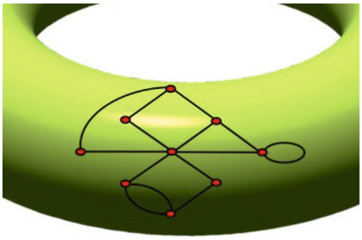
Can you tell which graph corresponds to which Platonic solid?

## The Euler Characteristic

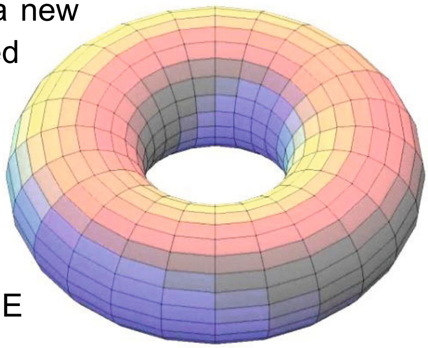
We have seen that Euler's formula works in at least three settings: the plane, the sphere, and the Platonic solids. We have described natural connections between these three settings.

Are there any settings in which Euler's formula does NOT work? Try this. Instead of drawing your connected graph on a plane or a balloon, draw it on an inner tube – the kind used to float down a lazy river. Or draw it on a "double inner tube" for couples who like to float together or on a "triple inner tube."

Use a permanent marker, but draw your graph when the water park attendant is not looking. Should you expect the expression V + F – E to equal 2 as before? That depends on how you draw



the graph. If you use only a small part of the rubber surface, like the illustration on the left, then you will get the same V, E, and F counts that you got when you drew that same graph on a piece of paper, so V + F – E = 2 as before. But the largest face of this graph is very peculiar. Unlike the other faces, its shape could not be formed by deforming (bending or stretching) a rubber polygon.



Let us start over and draw a new graph whose faces are all shaped like deformed polygons. For example, the faces of the graph on the right are all deformed squares. We carefully counted F = 576, V = 576 and E = 1,152. Therefore, the expression V + F – E equals zero (not 2).

In fact, the expression V + F – E equals zero for ANY connected graph embedded on an inner tube, no matter how complicated, as long as its faces are all deformed polygons. This is not much of a restriction. It is OK, for example, if some of its faces are triangles, while others are squares. The ones that are squares need not be perfect squares, but may be deformed (stretched and curved) rubber squares, like those on the inner tube graph pictured above.

It gets better. Any such graph on the "double inner tube" will satisfy V + F – E = –2. Any such graph on the "triple inner tube" will satisfy V + F – E = –4. These observations are the beginning of an entire field of mathematics called topology. Different surfaces can be distinguished by their value of V + F – E. This value is called the Euler characteristic of the surface. Thus, the Euler characteristic of the sphere equals 2, of the inner tube equals 0, of the double inner tube equals –2, and of the triple inner tube equals –4. Do you see the pattern? The Euler characteristic of a quadruple inner tube equals –6, and so on.

The Euler characteristic is unaffected by bending and stretching. You could stretch your spherical balloon into an egg shape, and this would not change its Euler characteristic because it would not change the V, E, or F counts for a graph drawn on the balloon. Thus, the Euler characteristic of a surface measures some essential quality of its shape that is unaffected by bending and stretching. Surfaces with different Euler characteristics cannot be bent or stretched into each other. If you meet a friend while floating down a lazy river, you cannot make room for her by stretching your single inner tube into a double, because the single and double tubes have different Euler characteristics.

## An Algebraic Proof that There Are Only Five Platonic Solids

Let us summarize all of the relationships that we have discovered among the numbers in our Platonic solid table, which is reprinted here:

| | V | E | F | S | C |
|---|---|---|---|---|---|
| Tetrahedron | 4 | 6 | 4 | 3 | 3 |
| Cube | 8 | 12 | 6 | 4 | 3 |
| Octahedron | 6 | 12 | 8 | 3 | 4 |
| Dodecahedron | 20 | 30 | 12 | 5 | 3 |
| Icosahedron | 12 | 30 | 20 | 3 | 5 |

First, we compared a Platonic solid to its dual: they have the same number of edges, their vertex and face counts are exchanged, and their “S” and “C” values are exchanged. These relationships are all explained by the geometric process of duality.

In addition, we discovered that the following four relationships are valid for each row of the table:

**(1)** $F \times S = 2 \times E$ **(2)** $V \times C = 2 \times E$ **(3)** $V + F - E = 2$
**(4)** $S \geq 3$ and $C \geq 3$.

We will prove now that the five rows of our Platonic solid table are the ONLY solutions to this system of four equations.

**ALGEBRA LEMMA:** There are only five ways to choose positive integers {V, E, F, S, C} so that equations **(1)**–**(4)** are satisfied; namely, the five rows of the Platonic solid table.

In other words, if you gave the four equations to a friend who had never heard of Platonic solids and who did not know what the variable stood for, your friend could, using only algebra, conclude that there are only five solutions to the system of equations. Here is how your friend would do it.

**PROOF:** Solving **(1)** for F gives F = (2 × E)/S. Solving **(2)** for V gives V = (2 × E)/C. Substituting these expressions for F and V into **(3)** gives: (2 × E)/C + (2 × E)/S – E = 2, which is the same thing as E × (2/C + 2/S – 1) = 2. In particular, we learn that:

(2/C + 2/S – 1) is positive.

Both S and C are at least 3. If either S or C were larger than 5, then (2/C + 2/S – 1) would NOT be positive. Therefore S = 3, 4, or 5 and C = 3, 4, or 5. If C and S were both 5, or both 4, or if one of them were 5 and the other were 4, then (2/C + 2/S – 1) would not be positive. Here are the only remaining possibilities:

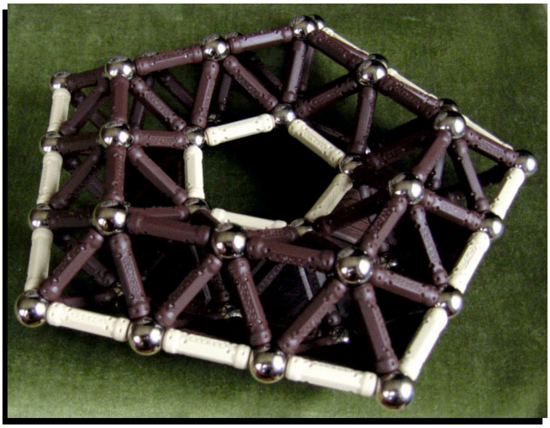
S = 3 and C = 3 (the tetrahedron), S = 3 and C = 4 (the octahedron), S = 3 and C = 5 (the icosahedron), S = 4 and C = 3 (the cube), S = 5 and C = 3 (the dodecahedron).

Once you know the values of S and C, you can use the equation "E×(2/C + 2/S – 1) = 2" to find the value of E, then use **(2)** to find the value of V and then use **(1)** to find the value of F……………□

It is fascinating that we can learn so much using only algebra. But how does The Algebra Lemma advance our study of the Platonic solids? A "lemma" means a mathematical fact whose primary purpose is to help prove a more important theorem. What important theorem will the Algebra Lemma help us prove?

For one thing, the Algebra Lemma provides an alternative way to prove that there are only five Platonic solids. Here is the idea. Suppose that your cousin Karl claims he discovered a 6th Platonic solid. We know that Karl's new solid must satisfy equations **(1)**–**(4)**. The Algebra Lemma tells us that Karl's solid must have the same values for {V, E, F, S, C} as one of the five familiar Platonic solids. But this means that Karl's solid is the same as one of these five familiar Platonic solids – it is not really new.

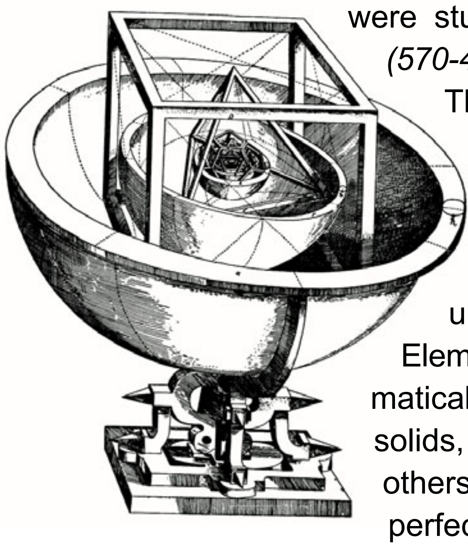
Most people regard this new proof as better than our previous cardboard-and-scissors proof. You can't argue with algebra. Nevertheless, it is important to confess that the new proof has its own shortcomings. For example, how do we know that Karl's solid satisfies Euler's formula? What if it has an Euler characteristic different from 2? The construction pictured here has an Euler characteristic zero. It is not a Platonic solid because different vertices have different numbers of triangular faces meeting at them. In fact, seven triangles meet at each vertex along the inner white rim, which is a nice illustration of overcrowding. But maybe Karl can figure out how to construct an inner tube shape like this that actually is a Platonic solid. A proof is required to rule out this possibility. We are not trying to cast doubt on the theorem – it is true that there are exactly five Platonic solids – we are just acknowledging that the proofs which we have discussed so far are not quite complete.



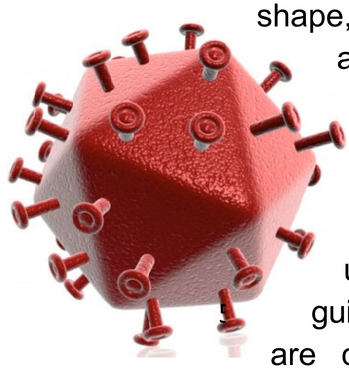
Not a Platonic Solid.
Photo by Karl Horton, Flickr.com

## The Platonic Solids Through the Ages

Human fascination with the Platonic solids extends back thousands of years. These solids probably first appeared in the artwork of the Neolithic people of Scotland *(~1400 B.C.)*. They were studied extensively by Pythagoras *(570-495 B.C.)* and the ancient Greeks. They are named after Plato *(428-328 B.C.)*, who associated them with earth (cube), air (octahedron), water (icosahedron), fire (tetrahedron) and the universe (dodecahedron). Euclid's Elements *(~300 B.C.)* includes mathematical constructions for all five Platonic solids, and a proof that there are no others. There are only five of these perfect shapes, which elevated their importance in scientific and theological writings. These five shapes found their way into all manner of theories. They were claimed to represent the fundamental pieces out of which all matter is formed. In 1659, Kepler explained the motions of the known planets using a model of the solar system that was based on the five Platonic solids inscribed inside each other (illustrated above).


Kepler's model of the solar system

Today, scientists do not view the Platonic solids as directly relevant to the motions of the planets or the fundamental building blocks of matter. Nevertheless, these five solids maintain an important status within math and science. As we saw in Chap. 7, every bounded solid object (which is not essentially two-dimensional) has the same proper symmetry group as one of the five Platonic solids (or a sphere). Thus, the Platonic solids are models for the possible ways in which the solid objects around us, like molecules in chemistry and cell structures in biology, can be symmetric. Why did the HIV virus evolve an icosahedral shape, as illustrated on the left? A biologist's answer to this question is informed by the mathematical fact that there are so few possibilities. Does modern science still regard the Platonic solids as representing earth, air, water, fire, and the universe? Not really, but here is an intriguing update. Scientists and mathematicians are currently using data collected by the Hubble telescope to attempt to discover the shape of the universe. One of the top contenders is an abstract three-dimensional shape which is called "Poincare dodecahedral space" because its geometry is intimately related to the symmetry group of a dodecahedron. Perhaps the dodecahedron does represent the universe!

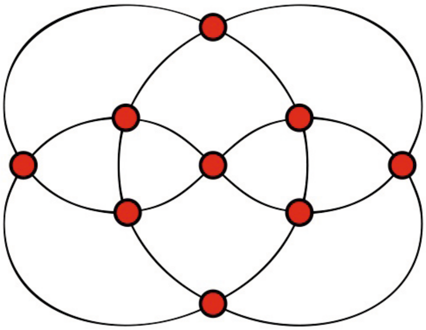


## Exercises

(1) For each dual pair of Platonic solids in our table, the values of "S" and "C" are exchanged. How does the geometric procedure of duality account for this relationship?

(2) The duality procedure can be applied to certain objects which are not Platonic solids. What solid results if you apply the duality procedure to a square-based pyramid?

(3) Count the vertices, edges, and faces of the planar graph below, and verify Euler's formula for this graph.



(4) You are asked to draw a connected planar graph with exactly 10 edges. Use Euler's formula to decide the largest and smallest possible number of faces that your graph could have. Then draw two graphs, one with the largest and one with the smallest possible number of faces.

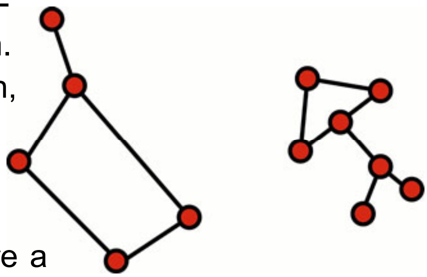
(5) If a connected planar graph has 12 vertices and 14 faces, how many edges does it have?

(6) We claimed that we carefully counted F = 576, V = 576 and E = 1,152 for the inner tube graph pictured in this chapter. That was a lie. Really, we just guessed the number of faces. We then knew that the graph had as many vertices as faces, and twice as many edges as faces. How did we know this?

COMMENT: *with these relationships,* $V + F - E = F + F - 2 \times F = 0$*, so the Euler characteristic comes out right even if our face count guess was wrong.*

(7) A soccer ball is made from 12 pentagon faces and 20 hexagon faces. Use this information to figure out how many vertices and edges it has, and then use this to verify that the Euler characteristic equals 2.

(8) The planar graph shown here is not connected; rather, it is built from two "connected components," with no bridge between them. Compute V, E, and F for this graph, and verify that V + F – E = 3. Prove that ANY planar graph with exactly two connected components will satisfy V + F – E = 3. Conjecture a formula for V + F – E in any planar graph with exactly *n* connected components. *HINT: What is the simplest such graph you can draw? From the simplest one, you can construct any other "one edge at a time."*



(9) For each Platonic solid, the number of proper symmetries equals F × S and also equals V × C. Explain why.

# 9. Symmetry and Optimization

Symmetry is beautiful, but that is not why there is so much of it in the world. Viruses evolved their icosahedral shape, not to be pretty, but because this shape optimizes performance. Bees evolved the behavior of building hexagonal honeycombs for functional, not esthetic reasons. The solution to an optimization problem is often highly symmetric. In other words, symmetric shapes are often the best shapes, and this principle helps account for their prevalence in nature.

## Minimal Surfaces

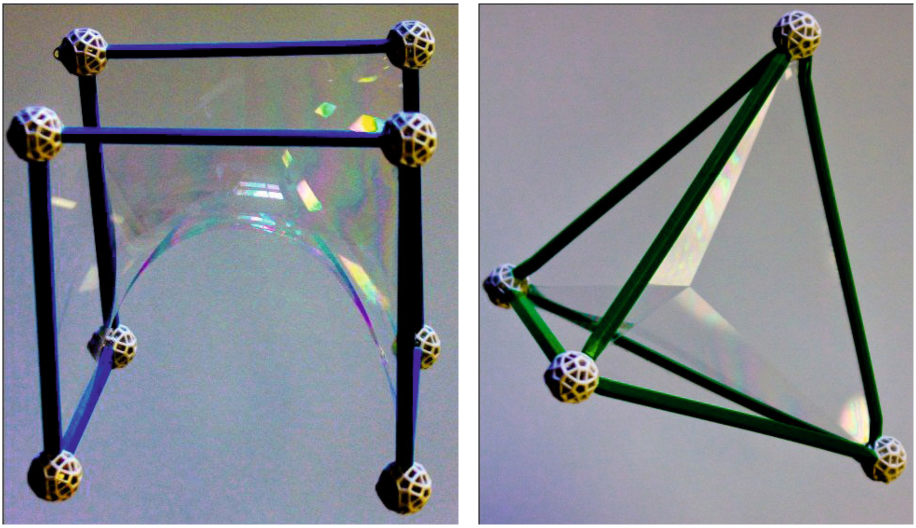
When you blow a bubble, it might be cigar-shaped at the instant it leaves the wand, but then it immediately snaps into a spherical shape. Why? What optimization problem is nature almost instantaneously solving here? Soap film is like elastic stretched taught. It wants to get smaller; that is, it wants to decrease its surface area. But it is forced to enclose your breath of air because surface tension prevents it from popping. Given that it must enclose a fixed volume of air, it finds the least-surface-area way to do so. The sphere is not just the best solution that the bubble can find; mathematicians have proven that it is the best among all conceivable shapes:

**THE BUBBLE THEOREM:** The sphere is the least-surface-area way to enclose a given volume.

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_9,

The sphere is also the most symmetric bounded three-dimensional shape, so at least in this case, the most symmetric shape is the optimum shape! As you read this chapter, keep in mind two principles: (1) The most symmetric shapes are often the best and (2) the solution to an optimization problem often has the same symmetries as the problem itself.

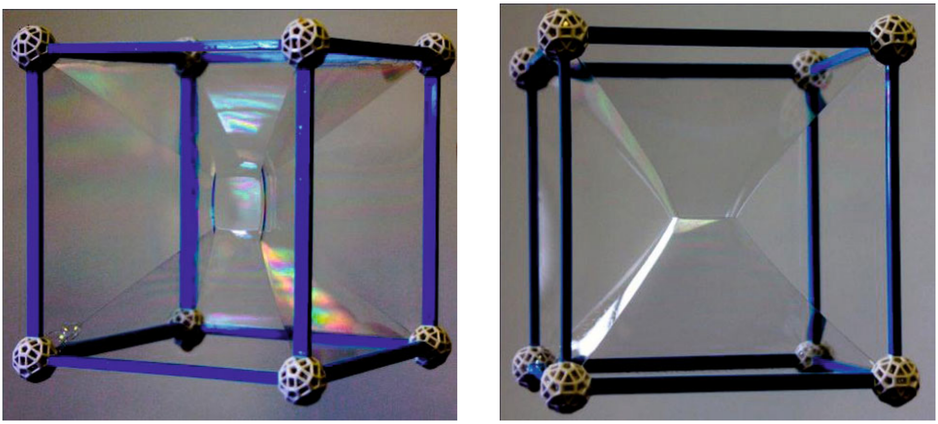
Soap solution can answer other optimization problems as well. If you dip a plastic frame into soap solution, nature finds the least-area surface stretching over that frame.



Least-area surfaces stretching over the saddle and tetrahedron frames

At the instant that the tetrahedron frame was removed from the bucket of soap solution, the soap film probably coated each of its four faces separately, but then it almost instantaneously snapped into the improved configuration pictured above. The surface could reduce its surface area further by letting go of the plastic frame, but chemical bonds prevent this. Given that it is compelled to cling to the plastic frame, it finds the least-area surface that does so. Notice that the soap film surface inside each frame pictured above has the same symmetries as the frame it clings to. This illustrates our second principle.

A least-area surface stretching over a cube frame is illustrated below, viewed from two different angles.



A least-area surface stretching over a cube frame, viewed from 2 angles

You might have expected the small square film in the center to collapse to a point, but a small square turns out to be better than a point. A good calculus student could verify this, but not nearly as quickly as the soap film figured it out.

The solution to an optimization problem *often*, but not always, has the same symmetries as the problem itself. Unlike the saddle and tetrahedron frames, the soap film in the cube frame does not have the same symmetries as the frame itself. This is possible because there are actually three soap film configurations that tie as least-surface-area winners, corresponding to whether the small square film in the center is parallel to the front, right, or top face of the cube frame. These three configurations are permuted by the symmetries of the cube (or you can actually shake the frame to make it switch between these three configurations). This is reminiscent of our observation in Chap. 7 that a largest possible cube that is able to fit inside a dodecahedron can fit in five different ways, and these five ways are permuted by the symmetries of the dodecahedron.

The study of least-area surfaces is currently an extremely active area of mathematics research. If you perform a web image search for "minimal surface," you will find a gallery of beautiful soap-film-like images, including bounded surfaces stretching over frames like those above and also unbounded surfaces extending indefinitely in all directions.

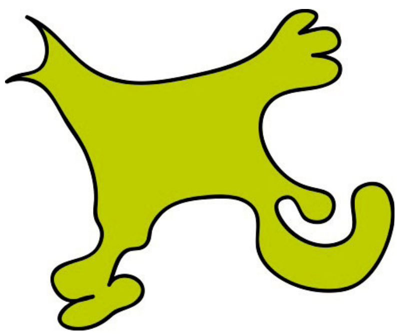
## The Circle Wins

There is a natural two-dimensional analog of The Bubble Theorem. Suppose that a farmer wants to enclose exactly 25 acres of grass land for his cows to graze on. How can he do so with the smallest possible length of fence? If you expect the most symmetric shape to be the best shape, then you will correctly guess the following theorem:

**THE CIRCLE THEOREM:** The circle is the least-perimeter way to enclose a given area in the plane.

Thus the circle, which is the most symmetric bounded object in the plane, is also the solution to the farmer's optimization problem. He should build a circular fence. The proof of this theorem is all about symmetry.

**SKETCH OF PROOF**. All of the farmers in the land competed in a contest to design the least-perimeter fence enclosing a given area (say 25 acres, although this number does not matter). Farmer Don won! His fence not only beat the other farmers' fences but also it beat all possible other fences. In other words, Farmer Don found the least-perimeter possible way to enclose the given area. We wish to prove that Farmer

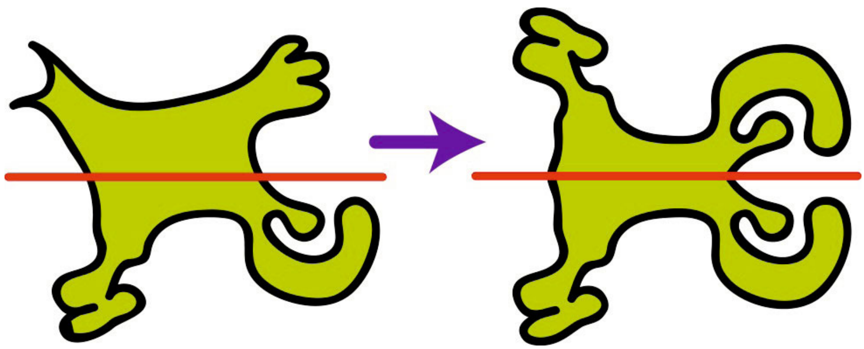


Farmer Don's winning fence might look like this.

Don's winning fence is a circle, but for all we know now, it could have a shape as crazy as the one illustrated above.

Consider the horizontal line that exactly divides the *area* of Farmer Don's fence in half. We claim that this line must also divide the *perimeter* of his fence in half. Why? Because if, say, the top had more perimeter than the bottom, then his fence would not have really been a winner – replacing the long top with the mirror-reflection of the short bottom would produce a fence that beats Farmer Don's original fence, contradicting our assumption that his original fence is the winner. In summary, because his fence is a winner, we know that the horizontal line that divides its *area* in half must also divide its *perimeter* in half. There are equal amounts of grass on each side *and* equal lengths of fence on each side.

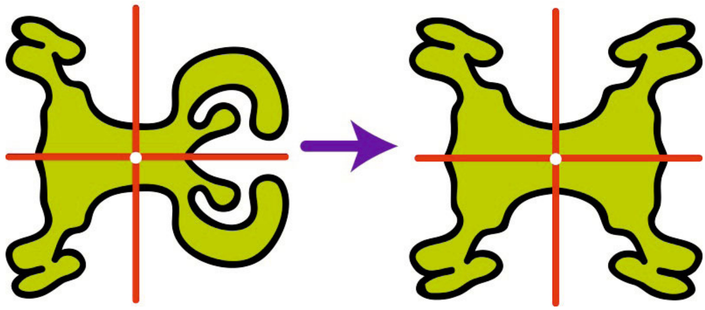
Now consider the new fence obtained by replacing the top half of Farmer Don's fence with the mirror reflection of the bottom half (over this horizontal halving-line), like this:



**Replacing the top half with the mirror reflection of the bottom yields a tied-winner!**

This new fence has the same area and same perimeter as the original, so it is a tied-winner! It ties with the original, but it is guaranteed to have at least two symmetries, whereas the original might have had only one (the identity).

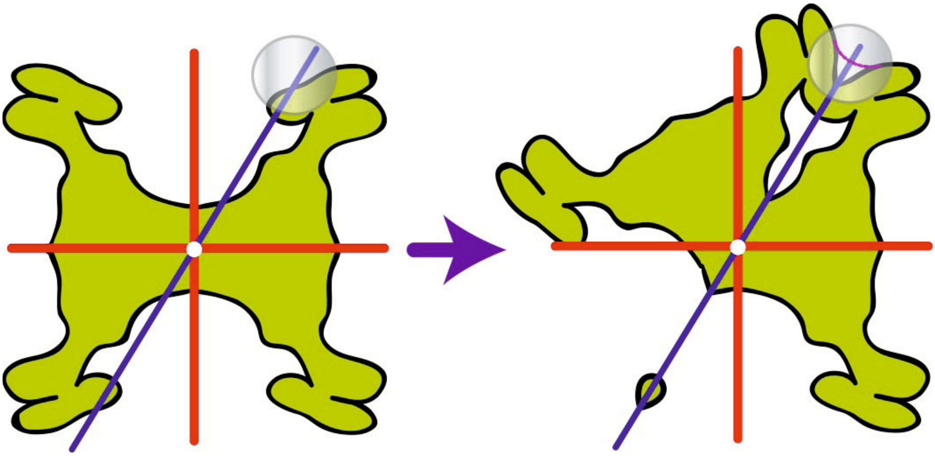
Next consider the *vertical* line that divides this new fence in half. As before, it must also divide the perimeter in half, so if we replace the right side with the mirror reflection of the left side, as pictured, then we obtain yet another tied winner.



Replacing the right half with the mirror reflection of the left yields another tied-winner!

This new tied winner is guaranteed to have at least four symmetries, namely **I**, **H**, **V** and $\mathbf{R_{180}}$. Notice that $\mathbf{R_{180}}$ is a symmetry because it is the composition of **H** and **V**, each of which is a symmetries. The white dot at which the horizontal and vertical lines cross is the center point of this new winner.

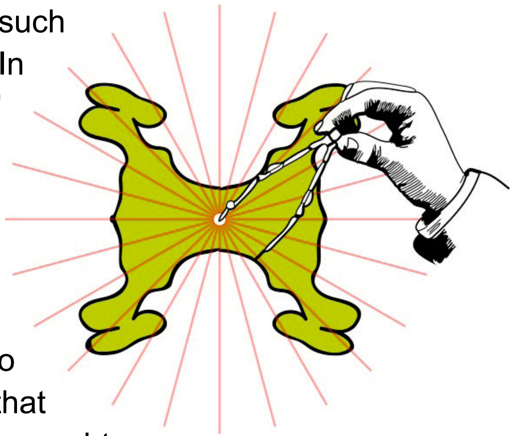
Observe that EVERY line through this white center point divides the perimeter and area of the new winner in half, simply because $\mathbf{R_{180}}$ exchanges the two sides of such a line. But this implies that every such line must meet the fence at right angles. Why is a non-right angle, like the angle at which the blue line pictured below meets the fence, impossible? Because the tied winner obtained by replacing one side of the blue line with the mirror reflection of the other side has an “innie-point” (two fence segments meeting at an angle pointing into the pasture side). But this is impossible; winners never have innie-points. The fence near any innie-point can be rounded off (as illustrated in purple) to yield a fence that is better on all counts: the purple modification encloses more area and has a smaller perimeter than the original.



If the blue line met the fence at a non-right angle, then replacing one side with the mirror image of the other produces something impossible: a winner with an innie-point.

In summary, starting with Farmer Don's winning fence, we built a tied winner with four symmetries, which is guaranteed to meet every line through its center point at a right angle.

Now let us brainstorm. Think of a fence shape that meets every radial line (meaning every line through its center point) at a right angle. How many different such shapes can you come up with? In fact, the only solution is a circle! To convince yourself of this, image tracing along the fence edge with a compass that's anchored at the center point, expanding and contracting the compass width as necessary to stay on the fence path. Notice that expanding the compass creates obtuse angles with the radial lines, while contracting creates acute angles. The only way to avoid acute and obtuse angles is to



never expand or contract your compass, which only happens when you are tracing a perfect circle.

Thus, the tied winner with at least four symmetries must be a perfect circle. This winner looks like four copies of the bottom-left quadrant of Farmer Don's original winner, so the original winner's bottom-left quadrant must be a quarter-circle. There is nothing special about the bottom-left. A slight modification to the above proof establishes that the bottom-right, top-left, and top-right quadrants are also quarter circles. Thus, Farmer Don's original winning fence must have been a circle........................□

That was a long proof! Hopefully it convinced you that the theorem is true. It is. Nevertheless, we must confess that more work is needed to make this proof precise and rigorous enough to satisfy mathematicians today. For example, the compass-tracing part is really a calculus problem which involves verifying that the polar coordinate function, $r(\theta)$, has zero derivative (if you do not know calculus, then please ignore that sentence). A more subtle issue is this: our proof showed only that a winning fence must be a circle. This conclusion is vacuous unless we can verify that a winning fence exists; in other words, that there exists a fence enclosing the given area that has smaller perimeter than any other fence (except for possible ties). This is true, and probably seems obvious, but it does require a separate proof. After all, there does not exist a *largest*-perimeter fence enclosing a given area.

It is not a bad thing that our proof was insufficiently rigorous. Beautiful new ideas are often discovered when we strive to put visual proofs and vague heuristic arguments on more solid and rigorous footing. In this case, proving that a winner exists lead to important ideas in the field of analysis. In any case, the details have been filled in, and we now know beyond doubt that the theorem is true: the circle wins!

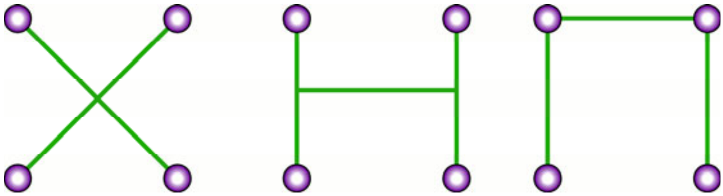
## Exercises

(1) Find the perimeter of the circle, square, and equilateral triangle enclosing area 25. Order these three shapes by increasing perimeter.

(2) Farmer Ann wants to fence off 25 acres of grass land along a straight river. She does not need to fence along the river's edge because her cows are afraid of water. Guess the least-perimeter fence shape.

(3) When two bubbles collide, they form a "double bubble" configuration which encloses and separates two (possibly different) volumes of air. Do actual soap bubbles find the least-surface-area way to do so, or do mathematicians know of a better solution that physical soap bubbles are unable to attain? Use online resources as needed to learn the answer to the double bubble problem.

(4) Four cities lie at the corners of a square. We wish to connect them by a network of roads so that it is possible to drive between any pair of cities. Which of the three road configurations pictured below uses the smallest total length of roadway? Can you invent a road configuration that beats all three of these configurations?



(۞5) The circle is the *most*-perimeter way to enclose a given area in the plane. This sentence is completely false, but my friend PK tried to prove it anyways. To prove it, he copied our proof of The Circle Theorem, replacing “least” with “most”, and making other such modifications as needed. For example, he easily changed the purple path near the innie-point to make it enclose *less* area and have *larger* perimeter than the original. What goes wrong with his attempt?

(۞6) The Circle Theorem is more commonly called *The Isoperimetric Theorem*, from the Greek word for “same perimeter.” This name is more appropriate for the following alternative version of the theorem: Among all curves with the same given perimeter, the circle encloses the most area. In other words, if a farmer has a fixed length of fence to work with, and he wishes to enclose the most possible area, he should build a circular fence. Explain why the two versions of this theorem are equivalent. *HINT: Explain why a counterexample to one version could be rescaled to become a counterexample to the other.*

# 10. What Is a Number?

By now you understand the importance of precise language. The history of mathematics is, among other things, a story about the invention of ever-more-precise language and techniques to explore the abstract ideas required to model the physical world. New ideas force us to look back and more precisely redefine our old vocabulary. This might sound like tedious backtracking, but it has been a driving force that has sparked some of the greatest breakthroughs in mathematics.

It is time now for us to backtrack. We began our story by defining a symmetry of an object as a rigid motion that leaves the object apparently unchanged. Now it is finally time to more precisely define the term "rigid motion." Historically, this was necessary to prove many of the theorems found in this book. Furthermore, since backtracking involves finding new ways to think about old ideas, it will lead us to unexpected discoveries and truths.

How far must we backtrack? Before deciding what a rigid motion of the plane or of space means, we must decide what "the plane" and "space" mean. Since the plane is made from pairs of numbers, $(x, y)$, and space from triples of numbers $(x, y, z)$, we must first decide what a number is. So let's start with numbers.

## Natural Numbers

In school, you first studied the natural numbers:

$$\mathbf{N} = \{1, 2, 3, 4, 5, 6, \ldots\} \text{ "the natural numbers"}$$

You learned that the most important natural numbers are the prime numbers: 2, 3, 5, 7, 11, 13, 17, 19, 23, and so on.

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_10,

**DEFINITION:** A prime number is a natural number greater than 1 that cannot be expressed as a product of two smaller natural numbers.

You probably learned that primes are the building blocks of all natural numbers. More precisely:

**THEOREM:** Every natural number greater than 1 is either prime or can be expressed in a unique way as a product of primes.

For example, the prime factorization of 300 is:

$$300 = 2 \times 2 \times 3 \times 5 \times 5.$$

We figured this out by breaking 300 down step-by-step until the pieces could not be further broken down…

…like this: $300 = 3 \times 100 = 3 \times 4 \times 25 = 3 \times 2 \times 2 \times 5 \times 5.$

…or like this: $300 = 10 \times 30 = 2 \times 5 \times 3 \times 10 = 2 \times 5 \times 3 \times 2 \times 5.$

The two answers above become the same if the primes are relisted in increasing order. That is what "in a unique way" means. Each natural number has *only one* prime factorization!

By the way, once you know the prime factorization of a number, it is easy to find the prime factorization of its square. For example, the square of 300 is $300^2 = 300 \times 300 = 90{,}000$, which has the following prime factorization:

$$90{,}000 = 300 \times 300 = 2 \times 2 \times 3 \times 5 \times 5 \times 2 \times 2 \times 3 \times 5 \times 5$$

$$\textit{(reorder)} = 2 \times 2 \times 2 \times 2 \times 3 \times 3 \times 5 \times 5 \times 5 \times 5.$$

Notice that $300^2$ has twice as many of each prime as 300 has in its prime factorization. In particular, the square of any natural number greater than 1 has an even number of occurrences of each prime in its prime factorization. This observation will be important in the next section. You'll see.

## Rational Numbers

After becoming acquainted with the natural numbers, you learned about the integers (which include zero and negatives):

**Z** = {…, –3, –2, –1, 0, 1, 2, 3, …} "the integers"

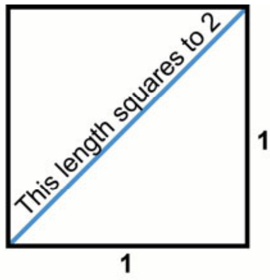
*(The German word for "integer" starts with Z)*. Next you studied rational numbers, which means fractions such as 3/5, 17/8, and –22/3:

**Q** = {all quotients "a/b" of integers with b≠0} "the rationals"

The terms "rational number", "ratio", "fraction," and "quotient" are roughly synonymous. Of course, some fractions are the same as others. For example, 2/3 = 4/6. If you were finding all solutions to an equation, like $3x=2$, you would not separately list 2/3 and 4/6, because these are the same number. Here is the general rule: a/b = c/d whenever ad = bc. Regard this comment as an addendum to our definition of **Q**. Thus, **Q** means the set of all quotients "a/b" of integers with b≠0, *with the understanding that certain quotients are really the same as others*.

If you studied mathematics 25 centuries ago in ancient Greece, you would have learned that all numbers are rational. What else could a number be? All physical matter was thought to be finitely dividable (that is, every object was thought to be made up of finitely many undividable building blocks), so rational numbers could describe the exact size of any portion of any object. Furthermore, the Greeks regarded rational numbers as divine gifts from their gods. When clues first appeared that other types of numbers existed, it was deemed blasphemous to pursue such thoughts.

These early Greek mathematicians developed number theory and geometry in parallel. They required numbers to represent not just portions of physical objects but also lengths in geometric constructions. For example, they knew that the diagonal of a 1-by-1 square has a length that squares to 2 (when multiplied times itself, the answer is 2). Today, we call this length “the square root of two,” denoted $\sqrt{2}$. They searched in vain for a fraction that squares to two, but their search led instead to a proof that no such fraction could ever be found.




**THEOREM:** There is no rational number that squares to 2.

Today, we say “$\sqrt{2}$ is a number which is not rational – an irrational number.” This sentence would have puzzled the early Greek mathematicians because, to them, “number” meant “rational number” (and also because they spoke Greek).

**PROOF**: Suppose your uncle Pete claims to have found two positive integers, p and q, such that the fraction p/q squares to exactly 2. That is, $(p/q)^2 = 2$. How do we know that Pete is mistaken? Let us explore the consequences of his claim and demonstrate that it leads to a contradiction. Re-write Pete’s claim with some simple algebra like this:

$$(p/q)^2 = 2 \leftrightarrow p^2/q^2 = 2 \leftrightarrow p^2 = 2 \times q^2$$

Now ask the question: how many 2s are in the prime factorizations of the left and the right sides of the equation $p^2 = 2 \times q^2$? The left side ($p^2$) has an even number of 2s because it has twice as many as p has. The right side ($2 \times q^2$) has an odd number of 2s because it has one more than twice as many as q has. But the prime factorization is unique. If the left and right sides really equaled each other, there would not be a difference between their prime factorizations. There is only one possible conclusion: Pete was mistaken. There is no rational number which squares to 2.......................................................□

When the early Greek mathematicians discovered this theorem, it was an abrupt challenge to their mathematical and religious belief systems, so they closely guarded this uncomfortable truth. Only a select few mathematicians were privy to the secret, and those who were caught sharing it with the uninitiated were executed!

## Real Numbers

Inevitably the secret got out, and it led to the invention of a number system including more than just fractions. What do rational numbers like 2/3 and lengths like $\sqrt{2}$ have in common, so that both can be incorporated into a more general concept of number? Today, we put them on equal footing by writing them as <u>decimal expressions</u>:

$$2/3 = 0.66666666\cdots \quad \text{and} \quad \sqrt{2} = 1.41421356237\cdots$$

In the decimal expression for 2/3, the symbol "$\cdots$" indicates a continuation of the pattern (an unending string of 6s). In the decimal expression for $\sqrt{2}$, the symbol "$\cdots$" just indicates some unending string of digits, whose pattern you might not understand. We are lead to a precise definition of a more general type of number:

**DEFINITION:** A <u>real number</u> means a "decimal expression"; that is, an expression formed from an integer followed by a decimal point followed by infinitely many digits. The set of all real numbers is denoted **R**.

A real number like 4/5 = 0.8 should be appended with an unending string of zeros ($4/5 = 0.800000\cdots$) so that is has infinitely many digits after its decimal point, as required by the definition. The purpose of this requirement is to put all real numbers on equal footing, so that pairs of them can be more easily added or multiplied.

Speaking of which, you might be embarrassed to learn that you do not really know how to add or multiply real numbers. For example, what is $2/3 + \sqrt{2}$? Familiar rules of addition require you to begin at the right-most digit and work left, but the above decimal expressions for 2/3 and $\sqrt{2}$ go on indefinitely to the right. There is no right-most digit at which to begin!

Precisely defining real number addition, multiplication, subtraction, and division involves technical intricacies that are beyond the scope of this book. We will settle for discussing only one important subtraction problem:

$$12.75000000000\cdots - 12.74999999999\cdots = ???$$

Can you guess the answer? The difference between the red and green number is clearly less than 12.75 – 12.74 = 0.01 and less than 12.750 – 12.749 = 0.001, and less than 12.7500 – 12.7499 = 0.0001, and so on. Since this difference is less than arbitrarily small numbers, the only reasonable guess is that the difference equals zero. In fact, zero is the correct answer! But if the difference between two numbers equals zero, then those numbers must equal each other:

$$12.74999999999\cdots = 12.75000000000\cdots$$

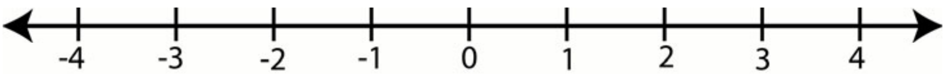
In case you have trouble believing this, a second purely algebraic way to see that these two numbers are the same is found in the exercises at the end of this section. The red and green decimal expressions are different ways of writing the same real number, just like 2/3 and 4/6 are different ways of writing the same rational number. Here is the general rule:

**REAL REDUNDANCY RULE:** A digit (other than 9) followed by an unending string of 9s can be replaced by the next larger digit followed by an unending string of 0s. There are no other redundancies among real numbers.

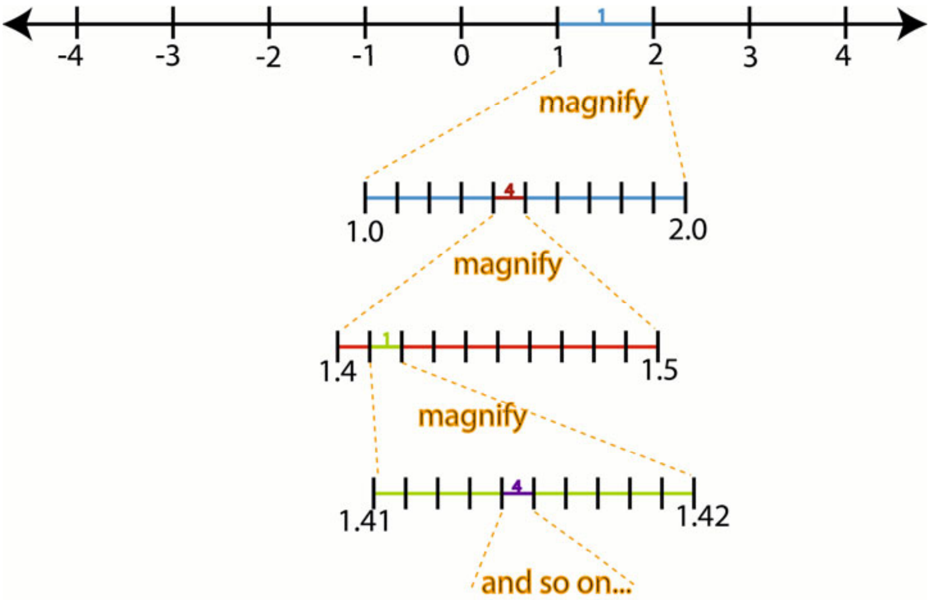
The strings of 0s and 9s are even allowed to straddle the decimal point. So if a car dealer quotes you $19,999.99⋯, then you will be making out your check for exactly 20,000.

The Real Redundancy rule is green-boxed because it is a definition, not a theorem. More precisely, the rule is an addendum to our definition of a real number. Thus, **R** means the set of all decimal expressions, *with the understanding that certain decimal expressions are really the same as others*.

Our “decimal expression” definition of a real number fits nicely with our visual intuition that the real numbers should represent all possible (positive and negative) lengths; that is, a real number should be a point on an idealized (infinitely thin) line, which we call the “real number line,” pictured here:




The digits of a real number tell us how to locate it on this number line. For example, to locate $\sqrt{2} = 1.41421356237\cdots$, the first digit “1” tells us to look between 1 and 2. If we divide this interval into 10 equal bins (numbered 0–9), the next digit “4” tells us which bin to look in. If we subdivide that bin into 10 equal sub-bins, the next digit “1” tells us which sub-bin to look in, and so on. Here is the picture:

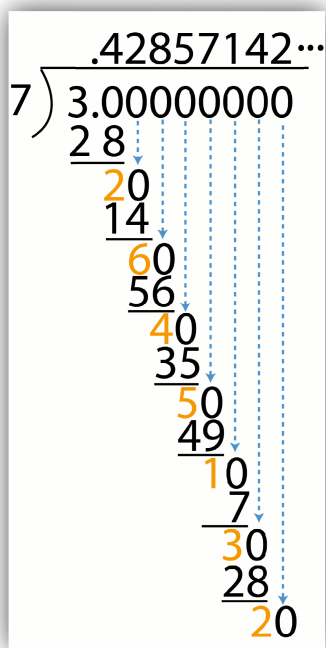



Each next digit provides a ten fold increase in the accuracy with which we know the number's location on the real number line. In fact, if you were asked to find $\sqrt{2}$ without a calculator, you would probably identify it digit-by-digit like in the above picture. First, $\sqrt{2}$ lies between 1 and 2 because 1 squares to less than 2, while 2 squares to more than 2. Next, $\sqrt{2}$ lies between 1.4 and 1.5 because 1.4 squares to less than 2, while 1.5 squares to more than 2, and so on.

## Which Real Numbers Are Rational?

Every rational number is a real number because long division can be used to convert any fraction into a decimal expression. For example:

3/7 = 0.428571428571428571$\cdots$ *(6-digit string repeats)*



The long division work of finding this answer is pictured on the right. Notice that the digits started repeating as soon as the orange numbers (the remainders) started repeating. Since there are only seven possibilities for this orange remainder (0–6), you knew in advance that it would repeat after not more than seven steps.

Could we work this long division problem backwards? That is, if $N = 0.428571428571428571\cdots$ had been given, could we have figure out that $N = 3/7$? Here is a clever trick. Since $N$ has a 6-digit repeating string, we will multiply it by 1,000,000 (which has six zeros), and then subtract N from 1,000,000 × N to remove the repeating string, like this:

$$\begin{array}{rl} 1{,}000{,}000 \times N = & 428{,}571.428571428571\cdots \\ N = & 0.428571428571\cdots \leftarrow \textit{subtract} \\ \hline 999999 \times N = & 428{,}571.000000000000\cdots \end{array}$$

We learn that $N = 428{,}571/999{,}999$ (which reduces to $N = 3/7$).

> **THEOREM:** A real number is rational precisely when its decimal expression is eventually repeating.

The term "eventually repeating" means that, possibly after some initial digits, the tail of the decimal expression is formed from an indefinite repetition of a single finite string of digits. For example, $12.3459214921492149214\cdots$ is eventually repeating, and so is the rational number $4/5 = 0.8$ because it gets rewritten as $0.80000\cdots$

**PROOF**: The previous examples show how to convert the fraction 3/7 into an eventually repeating decimal expression, and how to convert the eventually repeating decimal expression $0.428571428571428571\cdots$ into a fraction. Look back at these two examples and convince yourself that the same techniques would work to convert any fraction into an eventually repeating decimal expression, and any eventually repeating decimal expression into a fraction. If you are not yet convinced, try some practice problems from the exercises........................................................□

Because of this theorem, it is easy to find irrational numbers (real numbers that are not rational). Just design a decimal expression whose digits follow a pattern that's more complicated than "eventually repeating." For example, the following number is irrational:

$$N = 0.010010001000010000001\cdots \textit{ (pattern continues).}$$

Try to create your own examples of irrational numbers. Although it is easy to make up irrational numbers, it is often difficult to prove that particular numbers are irrational. For example, the famous numbers $\pi$ and $e$ are both irrational, but proving this is far beyond the scope of this book. It is still unknown whether $\pi^{\sqrt{2}}$ is rational or irrational, although most everyone expects it to be irrational.

## How Many Primes Are There?

The prime numbers played a very minimal role in our story about symmetry. In case you missed their relevance, here is a quick recap of our backtracking progress. To precisely define rigid motions of the plane and space, we must precisely define the plane and space, which requires us to precisely define real numbers, which requires an understanding that not all numbers are rational. The simplest provably irrational number is $\sqrt{2}$, and this proof requires an understanding of prime factorization.

In this final section, we embark on a short excursion for readers who desire to learn a few more fundamental and beautiful facts about prime numbers before returning to symmetry.

More specifically, we will explore the question: how many prime numbers are there? We take it as self-evident that there are infinitely many natural numbers. That is what the "…" meant when we wrote **N** = {1, 2, 3, 4, 5, 6, …}. But are there infinitely many primes? This is less obvious. After all, there are only finitely many Lego block shapes, but yet there are infinitely many different Lego constructions that could be built from unlimited supplies of them. It is similarly conceivable that there are only finitely many prime numbers out of which all of the infinitely many natural numbers can be built. This matter was settled by Euclid around 300 BC.

**EUCLID'S THEOREM:** There are infinitely many prime numbers.

**PROOF:** Imagine that your sister's boyfriend Andy insists that there are only finitely many prime numbers, and to prove it, he list all of them together on a piece of paper: $p_1, p_2, p_3, \ldots, p_n$. How do we know that Andy is wrong? No matter how lengthy his list, we will describe a strategy for identifying a prime number that is not on his list. Thus, no finite list of prime numbers could ever be complete.

Here is how we will identify a prime number that is missing from Andy's list (or from any finite list of prime numbers). First compute the result of multiplying all of Andy's primes together and adding 1. Call this number "L" because it is so large:

$$L = p_1 \times p_2 \times p_3 \times \ldots \times p_n + 1.$$

Notice that L is not on Andy's list because it is much larger than anything on his list. If L happens to be prime, then it is exactly what we seek: a prime number that is not on Andy's list. If L happens not to be prime, then any single prime number that appears in L's prime factorization is exactly what we seek: a prime number that is not on Andy's list. This is because none of Andy's primes divide evenly into L. In fact, L was custom built so that each of Andy's primes $p_1$, $p_2$, $p_3$, …, $p_n$ leaves a remainder 1 when you divide it into L. Thus, no finite list of prime numbers could ever be complete, which means there must be infinitely many prime numbers ………………………………………………□

Here is a quick example to help you understand the logic of the above proof. Suppose Andy's list is {2, 3, 5, 7, 11, 13}. We consider $L = 2 \times 3 \times 5 \times 7 \times 11 \times 13 + 1 = 30{,}031$. It turns out that L is not prime; its prime factorization is $30{,}031 = 59 \times 509$. The numbers 59 and 509 are primes missing from Andy's list.

So now we know that there are infinitely many primes, but we still might wonder how frequently occurring the prime numbers are among the natural numbers. Are primes in abundance, or are they a rare breed? Do most US citizens have a prime social security number, or very few? To warm up to questions like these, let us make a list of all natural numbers between 1 and 10, and highlight the primes, like this:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Notice that 40% of the numbers on this list are prime, which we will write as 0.40. Next we will go all the way to 100, like this:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100.

Notice that 25% of the numbers on this list are prime, which we will write as 0.25. As you travel further and further down the list of natural numbers, prime numbers occur in these proportions:

| N | Fraction of numbers up to N that are prime |
|---|---|
| 10 | 0.40 |
| 100 | 0.25 |
| 1,000 | 0.168 |
| 10,000 | 0.1229 |
| 100,000 | 0.09592 |
| 1,000,000 | 0.078498 |

The question is: what is the pattern in this table, and how does the pattern continue for larger and larger choices of *N*? Depending on how accurately you intend to answer this question, it is either difficult or extremely difficult or worth a million dollars.

The easiest answer is this: as *N* increases, the fraction approaches zero. The more difficult issue is specifying how quickly the fraction approaches zero. Here is a famous theorem, which was conjectured in the early 1800s and not successfully proven until 1896:

**THE PRIME NUMBER THEOREM:** The fraction of numbers up to *N* that are prime is approximately $\frac{1}{\text{twice the number of digits in N}}$.

This theorem is difficult to prove, but it is easy to use. For example, what fraction of numbers between 1 and 1,000,000 are prime? Since there are seven digits here, the answer is approximately $\frac{1}{14} \approx 0.0714$, which is close to the exact value from the table. Thus, about 7% of the numbers less than a million are prime.

Our version of the Prime Number Theorem above is not the most accurate version. For readers familiar with logarithms, we mention the more accurate and much more famous version, which says this: The fraction of numbers up to *N* that are prime is approximately $\frac{1}{\ln N}$. Our crude version above follows from this more accurate version because:

twice the number digits of N $\approx 2 \times \log N \approx \ln 10 \times \log N = \ln N$.

Our crude version sacrifices some accuracy (partly because 2 is not very close to $\ln 10 \approx 2.302\ldots$), but it has the advantage of being understandable to readers who are not familiar with logarithms.

Even the logarithm version of the Prime number theorem can be fine tuned and improved on. In fact, the quest to more precisely understand exactly how the prime numbers are distributed among the natural numbers is intertwined with some of the most difficult theorems in mathematics and also some of the most infamous unsolved problems. These include the Riemann Hypothesis, which today stands as the most famous unsolved math problem, with a one million dollar prize promised to the person who first solves it.

## Exercises

(1) Convert the following fractions into decimal expressions: 5/7, 23/21, 14/13.

(2) Convert the following decimal expressions into fractions:

$$10.792929292\cdots,\ 5.2003003003003\cdots,\ 0.34444\cdots$$

(3) Prove that $\sqrt{3}$ is an irrational number; in other words, there is no rational number that squares to 3. *Hint: Copy the proof in the chapter that $\sqrt{2}$ is an irrational number.*

(4) Prove that (5+$\sqrt{2}$)/7 is an irrational number. *HINT: If you could write it as a fraction, how could you use this to write $\sqrt{2}$ as a fraction?*

(5) Prove that $\sqrt{2}$+$\sqrt{3}$ is an irrational number. *HINT: Assume it is rational, which means $\sqrt{2}+\sqrt{3}= p/q$, then square both sides of this equation and find a contradiction.*

(6) Find a rational number between $M = 15.235950\cdots$ and $N = 15.237146\cdots$ *(to answer this, you do not need to know any more of the digits of M and N than are shown)*. Can you find infinitely many different rational numbers between $M$ and $N$? What about between $M = 15.236950\cdots$ and $N = 15.237146\cdots$?

(7) Find an irrational number between $M = 15.235950\cdots$ and $N = 15.237146\cdots$ *(to answer this, you do not need to know any more of the digits of M and N than are shown)*. Can you find infinitely many different irrational numbers between $M$ and $N$? What about between $M = 15.236950\cdots$ and $N = 15.237146\cdots$?

(8) Prove that there are infinitely many different rational numbers and infinitely many different irrational numbers between any pair of distinct real numbers.

(9) In your solution to the previous problem, why does your proof not work when the two real numbers are equal, like $M = 12.7499999\cdots$ and $N = 12.750000\cdots$?

(10) What is the smallest rational number larger than 0? If you do not think there is one, then explain why. What about the smallest irrational number larger than 0?

(⚙11) What can you say about the decimal expression for a fraction whose denominator has two digits? What is the longest the repeating string could possibly be? What is the longest possible string of initial digits before the repeating string begins?

(12) In this chapter, we saw that $N = 12.74999999999\ldots$ equals 12.75. Reprove this claim in a purely algebraic manner by converting $N$ to a fraction, and then showing that this fraction equals 12.75.

(13) To intelligently understand political news, you often need a frame of reference for comprehending big numbers. Filling in the following blanks (using internet resources as needed) may help you develop such a frame of reference:

A billion equals ________ million. The US population equals about _______ million. If the US government spends a billion dollars, this amount averages to about $_______ per citizen. A trillion equals ______ billion. If the government spends a trillion dollars, this amount averages to about $_______ per citizen. The US national debt is about $_______ trillion, which averages to about $_________ per citizen. The most recent national deficit equals about $_______ billion, which averages to about $________ per citizen. The difference between the meanings of the words "debt" and "deficit" is:__________________________________________________________.

(14) About what fraction of natural numbers between 1 and one trillion are prime? About how many natural numbers between 1 and one trillion are prime?

(15) About what fraction of US citizens have a prime social security number? About how many US citizens have a prime social security number?

(16) Suppose that X and Y are real numbers. For X, the third digit after its decimal point equals 5. For Y, the third digit after its decimal point equals 7. You do not know anything else about X or Y. In other words, X and Y look like this:

X = *.**5**** ⋯ and Y = *.**7****⋯

From this information, can you conclude that X and Y are DIFFERENT real numbers? What if X and Y instead look like this:

X = *.**9**** ⋯ and Y = *.**0****⋯

State a general rule for comparing two numbers which differ at a single decimal position.

(17) If X and Y are real numbers which differ at a single decimal position, and Y does NOT end in an infinite string of 9s or an infinite string of 0s, explain why X and Y must be *different* real numbers.

# 11. Cantor's Infinity

Thus far, we have introduced the following important sets of numbers:

**N** = {1, 2, 3, 4, 5, 6, …} "the natural numbers"

**Z** = {…, –3, –2, –1, 0, 1, 2, 3, …} "the integers"

**Q** = {all quotients "a/b" of integers with b ≠ 0} "the rationals"

**R** = {all real numbers} "the real numbers"

Which of these sets is the largest? You might respond that **R** is the largest because it contains the others. Or you might respond that they all have the same size, namely infinity. Until a little more than a century ago, mathematicians were content with the decision that every infinite set has the same size as every other infinite set. They were not right or wrong – this is simply what they meant by the phrase "same size."

**OLD-FASHIONED DEFINITION OF "SAME SIZE":** A pair of sets is said to have the same size if either (1) they are both finite and have the same number of members or (2) they are both infinite.

This definition probably seems reasonable, but you are about to learn a beautiful truth about infinity to which this definition blinds you. Mathematicians who used this definition did not understand their blind spot any more than the ancient Greek mathematicians understood the truths to which they were blinded when they defined "number" to mean "rational number." In the history of mathematical thought, this "infinity" blind spot was just as significant as the "number" blind spot, and its removal unleashed a rich world of fundamentally new ideas.

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_11,

## The Modern Meaning of "Same Size"

What else could the phrase "same size" possibly mean? To answer this question, let us think more carefully about how we compare the sizes of sets. When my niece was a toddler, I gave her ten candles and ten candle holders, and I asked her whether there were as many candles as candle holders. An adult would have separately counted the candles and holders and compared the answers, but my niece did not yet know how to count to ten. So instead, she simply placed one candle into each holder. Since the candles and holders matched up perfectly, she knew there were equal numbers of each.

If you are given two *infinite* sets and asked whether they have the same size, then your situation is very analogous to my niece's. You do not have the ability to separately count each set because you do not know how to "count to infinity." Your most reasonable solution is the one my niece used – you should try to find a one-to-one correspondence (a matching) between the members of the two sets. This idea is not child's play – it is so important, it will become our new meaning of "same size."

> **MODERN DEFINITION OF "SAME SIZE":** A pair of sets is said to have the <u>same size</u> if their members can be matched with a one-to-one correspondence.

It is time to forget the old-fashioned definition, and from now on, use only the modern definition. To decide whether two sets have the same size, your *only* job is to determine whether their members can be matched with a one-to-one correspondence. For example, to decide whether you have the same number of fingers as the stranger who you just met at the aquarium, you may not count and compare; rather, you must attempt a finger-to-finger matching as in the illustration below. It is often very natural to compare sizes by matching rather than counting. For example, in a truck load of new pairs of shoes headed to Payless, you know that the number of right shoes equals the number of left shoes without knowing how many of either are in the truck. Still, it is difficult to change old habits into new habits, so let’s practice.

We have the same number of fingers.

**EXAMPLE (A pair of finite sets):** How do we verify that the sets $S_1$ = {a, b, c, d, e} and $S_2$ = {A, B, C, D, E} have the same size? If you answered “they both have five members,” then you have not yet let go of the old-fashioned definition. From now on, the *only* way to confirm that two sets have the same size is to exhibit a one-to-one correspondence between their members, like this:

| a | b | c | d | e |
|---|---|---|---|---|
| ↕ | ↕ | ↕ | ↕ | ↕ |
| A | B | C | D | E |

**EXAMPLE (N and E):** How do we decide if **N** = {1, 2, 3, 4, …} (the set of all natural numbers) and **E** = {2, 4, 6, 8, …} (the set of all even natural numbers) have the same size? Here are some WRONG ANSWERS:

*“They have the same size because they are both infinite.”*

*“**N** is larger than **E** because **N** has all of **E**’s members plus more.”*

These answers are WRONG because they do NOT refer to the modern definition of “same size.” If you are tempted by these wrong responses, then you need to let go of your previous associations with the phrase “same size” and let the modern definition above become your ONLY meaning for this phrase. To answer this question, your ONLY job is to decide whether the member of **N** and **E** can be put into a one-to-one correspondence. After some trial and error, you will find that they can, like this:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | … |
| ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | |
| 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | … |

A one-to-one correspondence between the natural numbers and the even numbers

In case the pattern is not clear, we could describe it with a formula: $n \leftrightarrow 2n$. Do you see why this pattern is a one-to-one correspondence? For any even number you ask me about, I can find the natural number that matches with it. For example, if you ask me about 100, I report back that 50 matches with it. I will only ever have one choice for the number I report back, because this pattern never allows multiple natural numbers to match with the same even number. Thus, it is a one-to-one correspondence! We learn that **N** and **E** have the same size! How strange that an infinite set can have the same size as a subset of itself!

**EXAMPLE (N and Z):** Does **N** = {1, 2, 3, 4, …} have the same size as **Z** = {…, –3, –2, –1, 0, 1, 2, 3, …}? To answer this question, your ONLY job is to decide whether the member of **N** and **Z** can be put into a one-to-one correspondence. This seems difficult at first because the members of **Z** extend indefinitely in both the right (positive) and the left (negative) direction. But the following clever matching overcomes this difficulty:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | … |
| ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | ↕ | |
| 0 | 1 | –1 | 2 | –2 | 3 | –3 | 4 | –4 | … |

A one-to-one correspondence between the natural numbers and the integers

Do you see the pattern? The even natural numbers get matched with the positive integers, while the odd natural numbers get matched with the negative integers. In case the pattern is not clear, we could clarify it using a formula:

(even n) ↔ n/2, (odd n) ↔ –(n–1)/2

Thus, **N** and **Z** have the same size!

Finding a one-to-one correspondence can require cleverness and persistence. In the previous example, you might have first tried the matching n ↔ n, but then realized that this matching misses all of the negative members of **Z**. But just because one attempted matching fails, you can NOT conclude that the sets have different sizes – a cleverer attempt might still succeed.

**DEFINITION:** An infinite set is called countable if it has the same size as **N** (the set of natural numbers).

In the previous two examples, we concluded that **E** and **Z** are both countable. In general, to prove that an infinite set is countable, you must match its member with the natural numbers. That is, you must decide which 1st member of your set matches with 1, which 2nd member matches with 2, which 3rd member matches with 3, and so on. For **E** and **Z**, our matching looked like this:

| | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | … |
|---|---|---|---|---|---|---|---|---|
| **E** | 2 | 4 | 6 | 8 | 10 | 12 | 14 | … |
| **Z** | 0 | 1 | –1 | 2 | –2 | 3 | –3 | … |

In summary, to prove that an infinite set is countable, we must find an infinite listing of its members, {1st, 2nd, 3rd, … }, which is organized so as to eventually include each member.

## Are the Rational Numbers Countable?

Does **Q** (the set of all rational numbers) have the same size as **N**? In other words, is **Q** a countable set? If you believe that **Q** is countable, then to prove it you must find an infinite listing {1st rational, 2nd rational, and 3rd rational,…} organized so that your list eventually includes each rational number. Here is a first attempted pattern:

1/1, 1/2, 1/3, 1/4, 1/5, …

This attempt fails because it only includes positive fractions whose numerators equal 1. Let us improve this attempt by squeezing in more numerators:

1/1,1/2, 1/3, 2/3, 1/4, ~~2/4~~, 3/4, 1/5, 2/5, 3/5, 4/5, …
↑ *removed because 2/4 = 1/2*

This attempt is better – it eventually includes all positive fractions whose numerators are smaller than their denominators. Next, let us squeeze in their reciprocals:

1/1, 1/2, 2/1, 1/3, 3/1, 2/3, 3/2, 1/4, 4/1, 3/4, 4/3,
1/5, 5/1, 2/5, 5/2, 3/5, 5/3, 4/5, 5/4, …

The pattern here is: 1st fraction from the previous list, then its reciprocal, then the 2nd, then its reciprocal, and so on. This new pattern is better still – it eventually includes all positive fractions. All that remains is to insert 0 at the front, and intersperse the negatives:

0, 1/1, –1/1, 1/2, –1/2, 2/1, –2/1, 1/3, –1/3, 3/1, –3/1,
2/3, –2/3, 3/2, –3/2, …

The pattern here is: zero, then the 1st fraction from the previous list, then its negative, then the 2nd, then its negative, and so on. Behold the power of trial and error! We just proved:

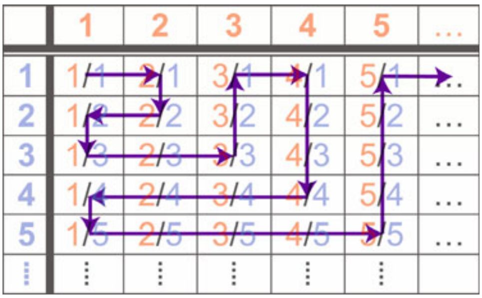
**THEOREM:** The set of rational numbers, **Q,** is countable.

The previous proof is perfectly valid, but the following alternative proof is also important to understand.

**ALTERNATIVE PROOF:** Here is an alternative method for listing all of the positive rational numbers. This means organizing them into an infinite list, but we first settle for organizing them into an infinite grid which (like a computer spreadsheet grid) has a top edge and left edge, but extends indefinitely down and right. The most natural arrangement is like this, with the column determining the numerator and the row determining the denominator:

| | **1** | **2** | **3** | **4** | **5** | … |
|---|---|---|---|---|---|---|
| **1** | 1/1 | 2/1 | 3/1 | 4/1 | 5/1 | … |
| **2** | 1/2 | 2/2 | 3/2 | 4/2 | 5/2 | … |
| **3** | 1/3 | 2/3 | 3/3 | 4/3 | 5/3 | … |
| **4** | 1/4 | 2/4 | 3/4 | 4/4 | 5/4 | … |
| **5** | 1/5 | 2/5 | 3/5 | 4/5 | 5/5 | … |
| **:** | : | : | : | : | : | |

Now we organize the cells of this infinite grid into an infinite list by meandering through the grid like this:



If we record the fractions we visit along this meandering purple path, and remove the redundant ones as we go, our list will begin like this:

1/1, 2/1, ~~2/2~~, 1/2, 1/3, 2/3, ~~3/3~~, 3/2, 3/1, 4/1, ~~4/2~~, 4/3, ~~4/4~~, 3/4,…

Now that we have successfully listed all of the positive rational numbers, we can insert zero at the front and intersperse the negatives as before……………………………………………………□

## Cantor's Theorem

Our next goal is to decide whether **R** (the set of all real numbers) is countable. To appreciate the question, try to construct an infinite listing {1st real, 2nd real, 3rd real, …}. You might start with a listing of the rational numbers and then insert some famous irrational numbers like $\pi$ and $\sqrt{2}$ at the front of your list. But what about the less famous irrationals, like the ones you made up yourself in the last chapter? The more you add to your list, the more you discover is missing. Are there too many real numbers to squeeze into a single infinite list? The answer to this difficult question was discovered by Georg Cantor around 1872.

**CANTOR'S THEOREM:** The set of real numbers, **R,** is NOT countable (so we call it <u>uncountable</u>).

I know lots of ways to construct an infinite listing of real numbers that fails to include them all. But this does not prove Cantor's theorem, since someone cleverer than me might someday succeed in including them all. To prove his theorem, Cantor had to show that NO listing, no matter how cleverly constructed, could ever succeed in including all real numbers. In other words, he had to prove that every attempted listing is doomed in advance. Here is how he did it:

**PROOF:** We will prove that any listing of real numbers is incomplete. No matter how scrupulously the list was organized, some real numbers were definitely left off. More precisely, we will describe a concrete procedure for identifying a real number that is missing from any given listing of real numbers.

Imagine a listing of real numbers. Maybe it was created by your Aunt Clair, who tried her best to include all of the real numbers on her list. Maybe it begins like this:

| | | | |
|---|---|---|---|
| 1st | ↔ | 3.1415926635⋯ | ($\pi$) |
| 2nd | ↔ | 0.3333333333⋯ | (1/3) |
| 3rd | ↔ | 1.41421356237⋯ | ($\sqrt{2}$) |
| 4th | ↔ | 256655643.0000000000⋯ | (Aunt Clair's SSN) |
| 5th | ↔ | 509.73737373737⋯ | (Her favorite number) |
| 6th | ↔ | 5.04749726737⋯ | ($\pi^{\sqrt{2}}$) |

Here is a concrete procedure for identifying a real number that is missing from the list. We will call this missing number M. It will lie between 0 and 1, so it will have the form

$$M = 0.d_1d_2d_3d_4d_5d_6d_7d_8\cdots$$

where each $d_n$ is a digit (0-9). How should we choose these digits to insure that M is NOT on the list? The answer is ingenious, and is hinted at by the red digits in Aunt Clair's list. Here it is: Choose M's first digit, $d_1$, to be anything other than the first digit (after the decimal point) of the first number on the list. This insures that M is different from the 1st number on the list, since it has a different first digit. Choose M's second digit, $d_2$, to be anything other than the second digit of the second number on the list. This insures that M is different from the 2nd number on the list, since it has a different second digit. Do you see the idea? Chose M's nth digit, $d_n$, to be anything other than the *n*th digit of the *n*th number on the list, which insures that M is different from the *n*th number on the list, since it has a different *n*th digit.

In the Aunt Clair example, the red diagonal includes the numbers {1, 3, 4, 0, 7, 7, …}, so we must choose

$$M = 0.(\text{not } 1)(\text{not } 3)(\text{not } 4)(\text{not } 0)(\text{not } 7)(\text{not } 7)\cdots.$$

This leaves us a lot of freedom. M = 0.258163··· works fine, as would many other choices. With each digit, there are ten choices (0–9), and only one choice is disallowed, which still leaves us nine options. To be on the safe side, we will also avoid 0s and 9s, which still leaves at least seven options for each digit. See Exercise #17 from Chap. 10 to understand the reason for avoiding 0s and 9s.

In summary, we can use this diagonal procedure to build a real number, M, which is missing from any given listing of real numbers. Therefore, no listing of real numbers could possibly be complete. Thus, the real numbers could never all be arranged into a single list – they are uncountable....................................□

Cantor's Theorem says that, in a very precise sense, the infinite sets **N** and **R** do NOT have the same size. Thus, the modern definition of "same size" leads to this truth: not all infinite sets have the same size – some are genuinely larger than others! This is a surprising and remarkable phenomenon. In popular writing, it is described with phrases like "different sizes of infinity" or "more infinite than infinity."

Georg Cantor

During Cantor's life, his work was criticized by theologians who considered it a challenge to the notion of God as the one and only infinite and also by mathematicians who were uncomfortable with his counterintuitive conclusions. But in the end, you can't argue with a solid proof. Cantor's conclusions were eventually accepted, causing a paradigm shift in the way mathematicians thought about fundamental concepts such as numbers and sets. The famous mathematician David Hilbert predicted the long-lasting importance of Cantor's work when wrote: "No one shall expel us from the Paradise which Cantor has created."

## Exercises

(1) Your brother in law has a list of four 4-digit numbers, but his handwriting is so poor, you can only make out one digit of each: 3***, *7**, **9*, ***0. Without knowing any more information, can you find a 4-digit number that is not on your brother in law's list? How is this problem related to Cantor's proof?

(2) Decide whether these sets have the same size: $S_1$ = {all odd natural numbers} and $S_2$ = {all natural numbers greater than 9}.

(3) Is it possible to list all of the positive rational numbers in increasing order?

(۞4) If a spreadsheet grid extended indefinitely left, right, up and down, would its cells be countable? *Hint: Can you find a path that meanders through all of its cells?*

(۞5) Prove that the set of irrational numbers is uncountable. *HINT: If it were countable, then interspersing a listing of all irrational numbers with a listing of all rational numbers would produce a listing of all real numbers.*

(۞6) If you replace the word "real" with "rational" throughout the proof of Cantor's Theorem, you get a faulty proof that the rational numbers are uncountable. Why is it faulty? Which step is incorrect?

(7) Is the set of all real numbers between 0 and 1 countable or uncountable? Why?

(8) Is the set of all real numbers whose decimal expressions contain no 5s countable or uncountable? Why?

(9) Is the set of all pairs of natural numbers countable or uncountable? Why?

(10) You have infinitely many piles, one for each natural number. The first pile has one marble, the second has 2 marbles, the third has 3 marbles, and so on. Is the total number of marbles a countable set?

(11) You have infinitely many piles – one for each natural number. Each pile has infinitely many marbles – one for each natural number. Is the total number of marbles a countable set?

(12) If you remove one member from an infinite set, will the new set always have the same size as the original set? *Hint: first consider the case when the original set is countable.*

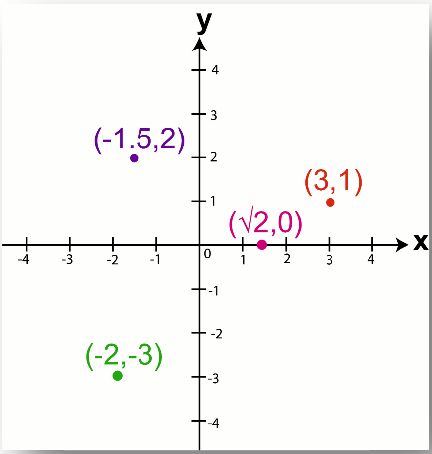
(13) Which of the following objects have countable symmetry groups: a circle in the plane, a sphere in space, the O border pattern. Explain your answers.

# 12. Euclidean Space

This book is about two-dimensional objects in the plane and three-dimensional objects in space, and the ways in which such objects can be symmetric. But what do the terms “plane” and “space” mean? To solidify and the foundation on which our theory of symmetry rests, we will now precisely define the words “plane”, “space,” and “rigid motions” thereof. Let’s begin.

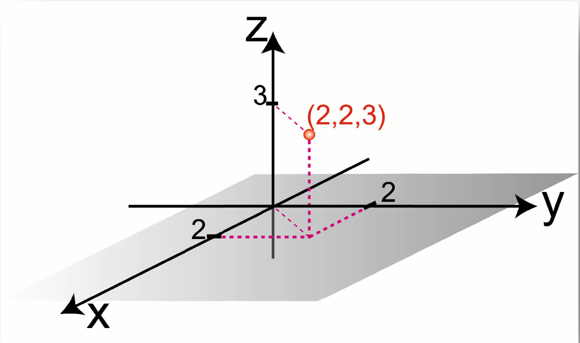
**DEFINITION:** <u>*n*-dimensional Euclidean space</u>, denoted $\mathbf{R}^n$, means the set of all ordered *n*-tuples of real numbers. $\mathbf{R}^2$ is called <u>the plane</u>, and $\mathbf{R}^3$ is called <u>space</u>.

The plane, $\mathbf{R}^2$, is thus the set of all ordered 2-tuples (pairs) of real numbers, like $(1,-7)$ and $(\sqrt{2},\pi)$. The illustration on the right shows several points of $\mathbf{R}^2$. A general point of $\mathbf{R}^2$ is denoted $(x,y)$. Notice how these two real numbers record the location of a point in the plane. The first number, $x$, describes its east–west position, while the second number, $y$, describes its north–south position.




Space, $\mathbf{R}^3$, means the set of all ordered 3-tuples (triples) of real numbers, like $(1,-7,5)$ and $(\sqrt{2},\pi,-18)$. A general point of $\mathbf{R}^3$ is denoted $(x,y,z)$. Notice how these three real numbers describe the location of a point in space. The first number, $x$,

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_12,

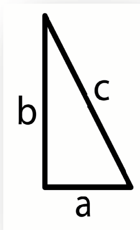



describes its east–west position, the second number, $y$, describes its north–south position, and the third number, $z$, describes its up–down position. The illustration on the left shows the location of the point $(2,2,3)$ in $\mathbf{R}^3$.

$\mathbf{R}^4$ means the set of all ordered 4-tuples $(x,y,z,w)$ of real numbers, for example $(\sqrt{2},\pi,-18,19)$. Although high dimensional Euclidean spaces like $\mathbf{R}^4$ cannot be visualized as well as $\mathbf{R}^2$ and $\mathbf{R}^3$, they are still of practical importance. For example, a food manufacturer who records his sales of seven different products each week is really recording a point of $\mathbf{R}^7$.

The purpose of this chapter is to explore the subtleties and intricacies of Euclidean spaces, particularly $\mathbf{R}^2$ and $\mathbf{R}^3$, and to define "rigid motions" of Euclidean spaces.

## The Pythagoran Theorem and Distance Formula

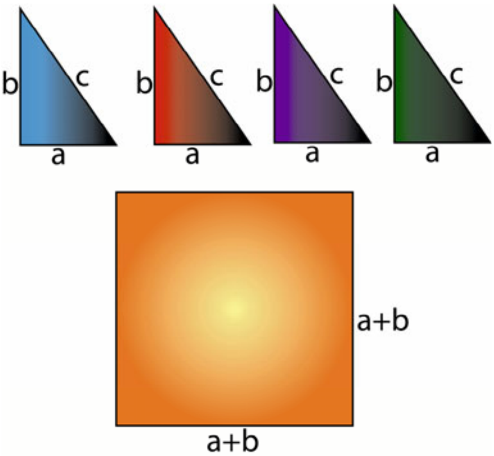



The most important fact about the plane, $\mathbf{R}^2$, was discovered by the Pythagoreans – the same group of ancient Greek mathematicians who discovered that $\sqrt{2}$ is irrational. It is a relationship between the lengths of the sides of a right triangle:

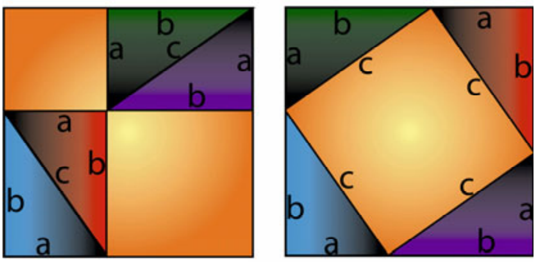
**THE PYTHAGOREAN THEOREM:** The lengths, a, b and c of the sides of a right triangle (listed in increasing order) satisfy the relationship: $a^2 + b^2 = c^2$.

For example, if a = 3 and b = 4, then $c^2 = 3^2 + 4^2 = 9 + 16 = 25$, so $c = \sqrt{25} = 5$.

**PROOF:** For the proof, you will need some supplies. Cut out four copies of the same right triangle. We will call its side lengths a, b, and c (in increasing order). Next, cut out a square whose side length is a + b. These supplies are pictured in the illustration on the right.




The key observation is that there are two very different patterns with which to arrange the four triangles onto the square.

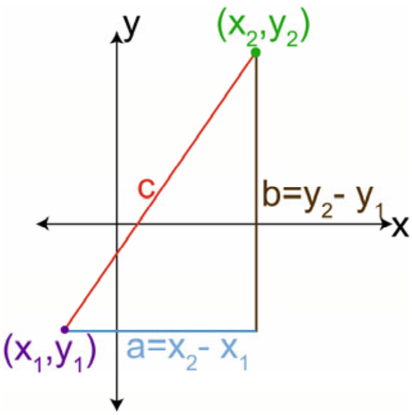



These two patterns are shown in the left illustration. For the first pattern, the uncovered orange area equals $a^2 + b^2$ (because it is made of two squares with side lengths a and b). For the second pattern, the uncovered orange area equals $c^2$ (because it is made of a single square with side length c). Since rearranging the pattern could not change the uncovered area, we conclude that $a^2 + b^2$ must equal $c^2$. If we had started with a fatter or narrower right triangle, think about why the two arrangements still work out, with corners meeting perfectly as in the illustration above, and with right angles at the corners of the orange c-by-c square.................................□

Why is the Pythagorean Theorem important? Because it allows us to measure distances between points in the plane:

**DISTANCE FORMULA FOR THE PLANE:** The distance between $(x_1, y_1)$ and $(x_2, y_2)$ in **R**$^2$ equals $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.

**PROOF**: This distance formula is nothing more than the Pythagorean Theorem in disguise. The distance between $(x_1, y_1)$ and $(x_2, y_2)$ equals the length, $c$, of the hypotenuse (the longest side) of the right triangle pictured in the illustration on the left. The Pythagorean Theorem says $c = \sqrt{a^2 + b^2}$, which is exactly the distance formula....................□




The distance formula in space is analogous:

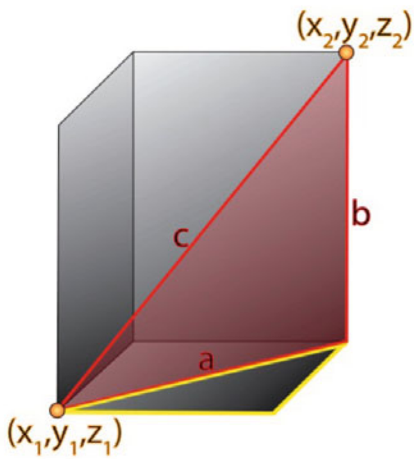
**DISTANCE FORMULA FOR SPACE:** The distance between $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$ in **R**$^3$ equals

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

**PROOF**: The proof involves a clever double application of the Pythagorean Theorem. In the illustration below, applying the Pythagorean Theorem to the red triangle gives that the distance between the two orange points in **R**$^3$ equals:

$$\text{Distance} = c = \sqrt{a^2 + b^2}.$$

Now simplify this expression for $c$ after substituting the following:




$$b = z_2 - z_1,\ a = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

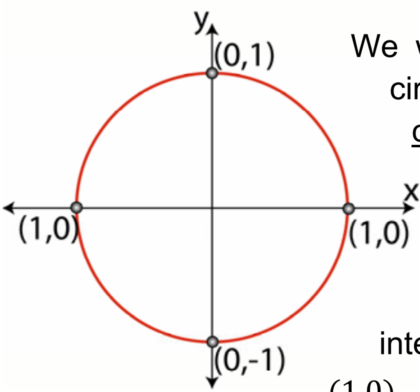
This expression for $a$ comes from applying the Pythagorean Theorem to the yellow triangle……………...………………………□

Our distance formulas generalize naturally to $\mathbf{R}^n$ in the most obvious manner. For example, the distance between the points $(x_1, y_1, z_1, w_1)$ and $(x_2, y_2, z_2, w_2)$ in $\mathbf{R}^4$ equals:

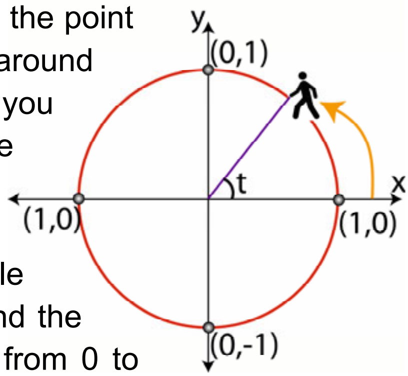
$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + (w_2 - w_1)^2}$$

and so on. We do not need to prove that this formula agrees with any previous meaning of "distance" in $\mathbf{R}^n$, because when $n > 3$, there is no previous meaning. We can, therefore, take this formula as our definition of the word "distance." For example, a food manufacturer who records his sales of seven different products each week might use the distance formula in $\mathbf{R}^7$ to compare this week's sales to last week's sales.

## Naming the Points on the Unit Circle




We will now take a closer look at the circle – more specifically, the unit circle in $\mathbf{R}^2$, which means the circle with radius 1 centered at the origin, $(0,0)$. How many points on the unit circle can you identify? The easy ones to identify are its intersections with the $x$ and $y$ axes: $(1,0)$, $(-1,0)$, $(0,1)$, and $(0,-1)$. Do you know the $x$ and $y$ coordinates of any other points on the unit circle? You could probably use the distance formula to find more, but to really understand the unit circle, it is not enough to just haphazardly name some points. What we really need to do is *parameterize* all of the unit circle's points in a systematic way.




Here is the idea. Beginning at the point $(1,0)$, you walk counterclockwise around the unit circle. In your left hand, you hold one end of a purple rope. The other end of the rope is tethered to $(0,0)$. Consider the angle that the positive $x$-axis makes with the purple rope. As you walk one full trip around the circle, this angle steadily increases from 0 to 360. We will call this angle your "angle-position" (because it determines where you are) and we will denote this angle-position with the letter "$t$."

The $x$ and $y$ coordinates of your position (when your angle-position equals some value of $t$ between 0 and 360) have special names:

**DEFINITION:** If you are at angle-position $t$ on the unit circle,

$cos(t)$ means your $x$-coordinate *(pronounced "the cosine of $t$").*

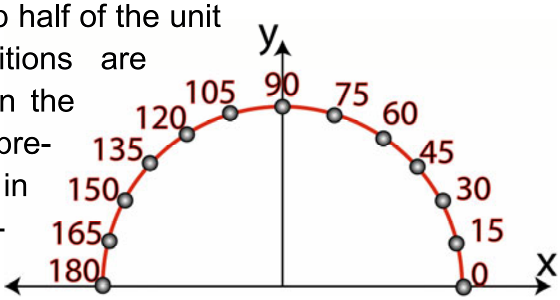
$sin(t)$ means your $y$-coordinate *(pronounced "the sine of $t$").*

Most calculators can compute the cosine and sine of any angle-position with dedicated buttons labeled "cos" and "sin." Here is a table showing the cosine and sine of all multiples of 15 between 0 and 180 (the values are rounded to two decimal places):

| $t$ | 0 | 15 | 30 | 45 | 60 | 75 | 90 | 105 | 120 | 135 | 150 | 165 | 180 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $cos(t)$ | 1 | .97 | .87 | .71 | .50 | .26 | 0 | -.26 | -.50 | -.71 | -.87 | -.97 | -1 |
| $sin(t)$ | 0 | .26 | .50 | .71 | .87 | .97 | 1 | .96 | .87 | .71 | .50 | .26 | 0 |

Cosine and Sine table

What this table really tells you is the $x$ and $y$ coordinates of the 13 points on the top half of the unit circle whose angle-positions are labeled in the diagram on the right. Could you have predicted which numbers in the table would be positive and which would be negative? Could you have predicted which would be close to zero and which would be close to 1?



Angle-positions around the unit circle

## The Dot Product and Perpendicularity

In this section, you will learn a simple method for testing perpendicularity. To get started, we must introduce the terms “norm” and “dot product.”

**DEFINITION:** If $p = (a, b, c)$ and $q = (x, y, z)$ are points of $\mathbf{R}^3$,

(1) The norm of $p$ (denoted $|p|$) means the distance from $p$ to the origin $(0,0,0)$, which is computed as: $|p| = \sqrt{a^2 + b^2 + c^2}$.

(2) The dot product of $p$ and $q$ (denoted $p \bullet q$) is defined as:

$$p \bullet q = ax + by + cz.$$

Norms and dot products of points in $\mathbf{R}^n$ are defined analogously, as these examples demonstrate.

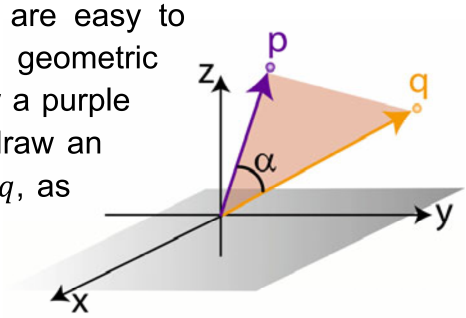
**Example in $\mathbf{R}^2$:** If $p = (3,4)$ and $q = (2,7)$ in $\mathbf{R}^2$, then the norm of $p$ is $|p| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$, and the dot product of $p$ and $q$ equals: $p \bullet q = 3 \times 2 + 4 \times 7 = 6 + 28 = 34$.

**Example in $\mathbf{R}^3$:** If $p = (2, -3, 5)$ and $q = (5, 7, -1)$, then the norm of $p$ is $|p| = \sqrt{2^2 + (-3)^2 + 5^2} = \sqrt{38} \approx 6.16$, and their dot is: $p \bullet q = 2 \times 5 + -3 \times 7 + 5 \times -1 = 10 - 21 - 5 = -16$.

**Example in $\mathbf{R}^4$:** If $p = (2,-3,4,1)$ and $q = (5,10,-1,0)$, then the norm of $p$ is $|p| = \sqrt{2^2 + (-3)^2 + 4^2 + 1^2} = \sqrt{30} \approx 5.48$, and their dot product is:

$$p \bullet q = 2 \times 5 + -3 \times 10 + 4 \times -1 + 1 \times 0 = -24.$$

Norms and dot products are easy to calculate, but what are their geometric meanings? To answer this, draw a purple arrow from the origin to $p$ and draw an orange arrow from the origin to $q$, as shown in the illustration on the right. In this picture, $|p|$ is simply the length of the purple arrow. Their dot product, $p \bullet q$, has a geometric meaning that depends on the angle (which we'll call $\alpha$) between the purple and orange arrows.




**THE MEANING OF THE DOT PRODUCT:** If $p$ and $q$ are points in $\mathbf{R}^n$, and $\alpha$ is the angle between the arrows pointing to them from the origin, then $p \bullet q = |p||q|cos(\alpha)$.

That is, the dot product can be found by multiplying these three numbers together: the norm of $p$, the norm of $q$, and the cosine of the angle between them. We will not prove this formula, but we will tell you why you should care about it.

Our main purpose for this dot product formula is the following quick and easy method for testing whether the purple and orange arrows are perpendicular:

**PERPENDICULARITY TEST:**

If $p \bullet q > 0$, then $\alpha$ is acute *(less than 90°)*.

If $p \bullet q = 0$, then $\alpha$ is right *(equal to 90°)*.

If $p \bullet q < 0$, then $\alpha$ is obtuse *(greater than 90°)*.

In particular, the orange and purple arrows are perpendicular exactly when the dot product equals zero. This perpendicularity test works because the cosine of an acute angle is positive, the cosine of a right angle is zero, and the cosine of an obtuse angle is negative. Think about why.

**EXAMPLE:** If $p = (2, -3, 5)$ and $q = (5, 7, -1)$ in $\mathbf{R}^3$, then their dot product is $p \bullet q = -16$, so the angle is obtuse.

**EXAMPLE:** If $p = (3,4)$ and $q = (2,7)$ in $\mathbf{R}^2$, then $p \bullet q = 34$, so the angle is acute.

**EXAMPLE:** If $p = (4,2)$ and $q = (-3,6)$ in $\mathbf{R}^2$, then their dot product is $p \bullet q = 4 \times -3 + 2 \times 6 = -12 + 12 = 0$, so the angle is right, which means that the arrows are perpendicular.

See how easy it is to determine whether two arrows are perpendicular? But what if you wish to know, not just whether the angle is acute or obtuse, but exactly what the angle equals? For this, just solve the “Meaning of the Dot Product” formula for $cos(\alpha)$, which yields:

$$cos(\alpha) = \frac{p \bullet q}{|p||q|}.$$

This tells you what the cosine of $\alpha$ equals. Then use the “$\cos^{-1}$” button on your calculator, which un-does the cosine function, to find what $\alpha$ equals. In the plane $\mathbf{R}^2$ and space $\mathbf{R}^3$, this answer is exactly the angle that you would measure with a protractor. In higher dimensional Euclidean spaces, protractors don’t make sense, so we simply take this answer as our definition of the word “angle” in higher dimensional Euclidean spaces.

## Using the Dot Product to Find a Lover or a Song

Here is an activity that might give you a better feeling for distances, dot products and angles. This activity will also help you understand the need to compute these quantities in high dimensional Euclidean spaces. But let's start with $\mathbf{R}^2$.

Write down your own personal point $(x, y)$ in $\mathbf{R}^2$ which represents $x$ = how much you love cats and $y$ = how much you love sushi, each on a scale from –5 (hate) to +5 (love). For example, my personal point is K = $(-3, 5)$ because I am allergic to cats and I love raw salmon. I named my point "K" because my first name is Kris. Next, ask a friend to write down his or her personal point. Draw your point and your friend's point in $\mathbf{R}^2$, and draw arrows to them from the origin. Next do three calculations:

(1) Calculate the dot product of your point and your friend's point. To what extent do you think that this dot product measures the compatibility of your interests? If the dot product is largely positive, does this mean that your interests are closely aligned with your friend's? If the dot product is largely negative, does this mean that the two of you have opposite interests? What does it mean if the dot product equals zero?

(2) Calculate the angle between the arrows. To what extent do you think that this angle measures the compatibility of your interests? Does an acute angle mean that your interests are closely aligned? Does an obtuse angle indicate opposite interests? What does a right angle indicate?

(3) Calculate the distance between your point and your friend's point. To what extent do you think that this distance measures the compatibility of your interests? If the distance is small, does this mean that your interests are closely aligned? If the distance is large, does this indicate that you have opposite interests?

Which quantity does the best job of measuring the compatibility of your interests: (1) the dot product, (2) the angle, or (3) the distance? There is no right answer – each strategy has advantages and disadvantages. But notice that ALL three quantities can be computed using dot products. Here is why:

**NORMS, DISTANCES, AND ANGLES CAN ALL BE COMPUTED USING DOT PRODUCTS:**

**(1)** $|p| = \sqrt{p \bullet p}$ *(the norm of a point is the square root of its dot product with itself)*.

**(2)** The distance between $p = (a, b, c)$ and $q = (x, y, z)$ equals

$$|p - q| = |(a - x, b - y, c - z)|$$
$$= \sqrt{(a - x, b - y, c - z) \bullet (a - x, b - y, c - z)}.$$

**(3)** $cos(\alpha) = \frac{p \bullet q}{|p||q|}$.

In summary, knowing how to compute dot products allows you to also compute norms, angles, and distances. All three strategies for measuring the compatibility of your interests are really just based on dot product calculations.

The previous activity might remind you of those dating websites that help single people find other single people with compatible interests. How does such a website match you up with your perfect future spouse? First, the site asks you about your interest level in more than just cats and sushi. Let us suppose the site asks you 50 questions, so that your personal point lies in $\mathbf{R}^{50}$. As discussed above, there are several reasonable strategies for measuring how “close” your personal point is to the personal point of someone else, like say Matt Damon or Penelope Cruz. The dot product is at the heart of all such strategies.

The administrators of dating websites are not the only folks who make money by calculating dot products in high dimensional Euclidean spaces. The creators of the Music Genome Project and Pandora Radio have also made a bundle.

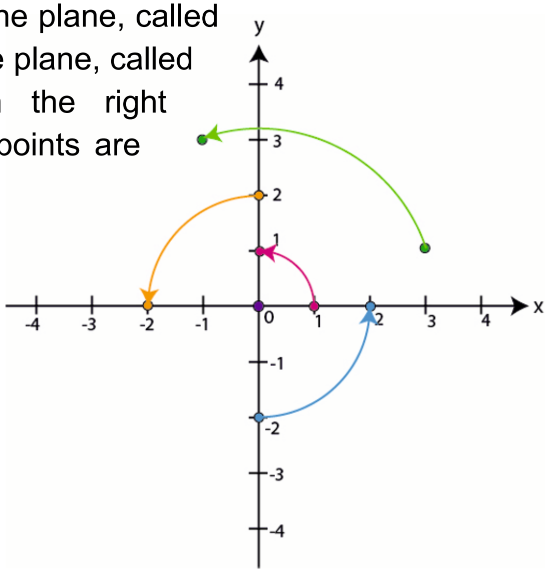
Their idea was to encode the musical essence of a song as a point in a high dimensional Euclidean space – as high as $\mathbf{R}^{400}$ for some music genres. The 400 numbers represent 400 musical attributes of the song: how twangy are the guitar solos, how angry are the lyrics, what gender is the lead vocalist, what is the song's time signature, and so on. Identifying the point in $\mathbf{R}^{400}$ that encodes one song takes a trained musician about a half an hour. They analyzed a large collection of songs and founded a music recommendation website based on their song-point database. A user of their website enters a song that he or she enjoys, and the website creates a personal radio station of similar songs. The user provides further information about his or her preferences by indicating approval or disapproval of each song played, and the station takes this added preference information into account in selecting future songs to play.

Pandora Radio keeps their algorithms secret, but it is fun to guess how the songs are selected. To find new songs similar to the songs that you like, their algorithm at core must measure the similarity of songs based on the "closeness" of the corresponding points in $\mathbf{R}^{400}$. As mentioned previously, the most natural measurements of closeness are all based on the dot product – in this case the dot product in $\mathbf{R}^{400}$. You might think of this as the "dot" in Pandora.com.

## What is a Rigid Motion?

We have been studying rigid motions since Chap. 1. At long last, we are finally equipped to explain precisely what this term means. The key idea is to regard a rigid motion of $\mathbf{R}^n$ as a function from $\mathbf{R}^n$ to $\mathbf{R}^n$. That is, a rigid motion is a rule or formula which allows one to determine the output point in $\mathbf{R}^n$ associated to each input point in $\mathbf{R}^n$. Visually, the output point represents where the motion "moves" the input point to.

For example, one of the first rigid motions of $\mathbf{R}^2$ that we studied was $\mathbf{R}_{90}$ – the 90° counterclockwise rotation of the plane about the origin. For practice, let us regard $\mathbf{R}_{90}$ as a function from $\mathbf{R}^2$ to $\mathbf{R}^2$. It inputs a point of the plane, called $p$, and it outputs a point of the plane, called $\mathbf{R}_{90}(p)$. The illustration on the right shows how several colored points are moved by this rotation:




$\mathbf{R}_{90}(0,0) \;=\; (0,0)$

$\mathbf{R}_{90}(1,0) \;=\; (0,1)$

$\mathbf{R}_{90}(3,1) \;=\; (-1,3)$

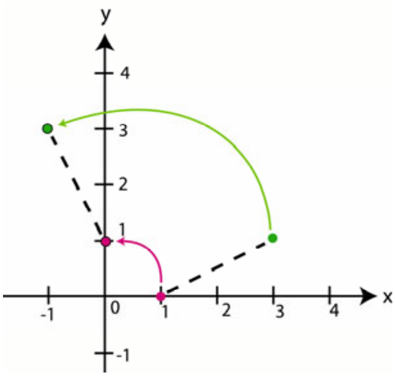
$\mathbf{R}_{90}(0,2) \;=\; (-2,0)$

$\mathbf{R}_{90}(0,-2) \;=\; (2,0)$

Do you see that pattern? You might guess that the pattern is this: $\mathbf{R}_{90}(x,y) \;=\; (-y,x)$. In the next chapter, we will prove that this formula is correct. Assuming this for now, this example helps us get used to thinking of a rigid motion as a function. Do you see how $\mathbf{R}_{90}$ is a definite rule that associates an output point to each input point? The rule is described by the formula $\mathbf{R}_{90}(x,y) \;=\; (-y,x)$. This formula empowers you to quickly determine $\mathbf{R}_{90}$'s effect on any input point you like. For example, $\mathbf{R}_{90}(37,55) \;=\; (-55,37)$. This is a whole new way to think about $\mathbf{R}_{90}$.

Translations are even easier than rotations to describe as functions. For example, let $\mathbf{T}_3$ denote the translation of the plane a distance 3 to the right. Considered as a function, $\mathbf{T}_3$ simply increases the *x*-coordinate of each point by 3. A general formula is $\mathbf{T}_3(x,y) \;=\; (x+3,y)$. For example, $\mathbf{T}_3(3,5) \;=\; (6,5)$ and similarly $\mathbf{T}_3(10,-3) \;=\; (13,-3)$.

When regarded as a function, the defining property of a rigid motion, **F**, is this: it preserves distances. This means that the distance between any pair of points is the same before and after the rigid motion is applied. That is what makes it rigid! In other words, the distance from $p$ to $q$ must equal the distance from $\mathbf{F}(p)$ to $\mathbf{F}(q)$, for any pair of points $p$ and $q$. For example, in the illustration on the right, the two dashed lines have the same length because $\mathbf{R_{90}}$ preserves the distance between the pink and green points.



$R_{90}$ preserves distances

At last, we can precisely define the vocabulary words upon which our entire study of symmetry has been built:

**DEFINITIONS:**

A rigid motion of $\mathbf{R^n}$ means a function, **F**, from $\mathbf{R^n}$ to $\mathbf{R^n}$ that preserves distances.

An object in $\mathbf{R^n}$ means a set of points in $\mathbf{R^n}$.

A symmetry of an object in $\mathbf{R^n}$ means a rigid motion, **F**, of $\mathbf{R^n}$ which moves the object onto itself; that is, $\mathbf{F}(p)$ is a point of the object whenever $p$ is a point of the object.

In truth, this definition of “object” only models single-colored or single-material objects. For example, we previously visualized a two-dimensional object as pained on the glass plane. If only white paint was used, then you may now more precisely regard this object as the set of points of $\mathbf{R^2}$ that have paint on them. But the above definitions must be modified to allow for multicolored objects. Think about how. Similarly, we previously visualized a three-dimensional object as sculpted, perhaps out of bronze, and surrounded by an infinite expanse of ice. You may now more precisely regard this object as the set of points of $\mathbf{R}^3$ where there is bronze rather than ice. But if multiple building materials were used, such as bronze, wood and paint, then you will need to think on your own about how to generalize the above definitions to incorporate this added generality.

It is difficult to picture an object in higher dimensional Euclidean spaces, but a picture is not always required. For example, the collection of songs written by Leonard Cohen is an object in $\mathbf{R}^{400}$. This object is unlikely to have any symmetries other than the identity.

## Two Exotic Examples

How well do the precise definitions of “object”, “rigid motion,” and “symmetry” agree with the intuitive feelings you have by now developed for these terms? It is important for the intuition and the rigor to be closely aligned, or at least to understand the ways in which they differ. Perhaps you previously thought of an object in $\mathbf{R}^2$ as something that you could paint or draw in the plane, and an object in $\mathbf{R}^3$ as something you could build out of plastic or wood or metal. However, our precise definition of the word “object” allows some things that can only be made with mathematical formulas not with paint or wood. Here are two examples.

**EXAMPLE: AN EXOTIC BORDER PATTERN:** When we classified the seven border patterns, we only considered border patterns that have a “smallest translation”. What was the purpose of that restriction? For one thing, it ruled out the $x$-axis, which is a border pattern that can be translated any amount right or left. But it also ruled out much more complicated border patterns. For example, consider the set of all rational numbers on the $x$-axis. In other words, the set of all point in $\mathbf{R}^2$ of the form $(x, 0)$, where $x$ is a rational number. This is a border pattern, but you would be hard pressed to draw an accurate picture of it. A translation right or left by any rational length is a symmetry of this border pattern because the sum of two rational numbers is rational, so the pattern is moved onto itself. A translation right or left by any irrational length is NOT a symmetry of this border pattern because a rational plus an irrational is irrational, so the pattern is not moved onto itself. This border pattern's symmetry group is very different from the symmetry groups of the seven border patterns in the classification theorem.

**EXAMPLE: AN EXOTIC BOUNDED OBJECT:** Look back at Da Vinci's classification of the possible symmetry groups of bounded objects in the plane. Notice that his theorem said nothing about bounded objects with infinite symmetry groups. Surprisingly, the circle is NOT the only such object. For example, consider the set of all points on the unit circle whose angle-positions are rational; that is, points that have the form $(cos(t), sin(t))$ for some rational number $t$. This is a bounded object in $\mathbf{R}^2$, although you would be hard pressed to draw an accurate picture of it. A rotation (centered at the origin) by any rational angle is a symmetry of this object, but a rotation by any irrational angle is not a symmetry of this object. To see why this is true, just notice that the rotation by angle $\alpha$ moves each point of this object according to this formula: $\mathbf{R}_\alpha(cos(t), sin(t)) = (cos(t+\alpha), sin(t+\alpha))$. So if $\alpha$ is rational, then $\mathbf{R}_\alpha$ moves this object onto itself, but if $\alpha$ is irrational, then it does not.

We mention the above examples in part to demonstrate the necessity of the fine print in the classification theorems from Chap. 4. We may not have anticipated examples like these, but we have no choice but to accept them. Surprising and counterintuitive examples play a crucial role in mathematics. The help delineate the true from the false, and studying them carefully can lead us to rich new understandings.

# Exercises

(1) In the proof of the Pythagorean Theorem, explain why the orange square with side length $c$ is really a square.

(2) If $p = (2,7)$ and $q = (3,-5)$ in **R**$^2$, find the following quantities: $|p|, |q|, p \bullet q$, the distance from $p$ to $q$, and the angle between the arrows from the origin to $p$ and $q$.

(3) If $p = (1,4,3)$ and $q = (-1,2,7)$ in **R**$^3$, find the following quantities: $|p|, |q|, p \bullet q$, the distance from $p$ to $q$, and the angle between the arrows from the origin to $p$ and $q$.

(4) In **R**$^5$, do the arrows from the origin to $(1,-2,3,-4,5)$ and $(1,0,-1,3,2)$ form an acute, obtuse or right angle?

(5) Let $p = (3,7)$. Show that $q_1 = (-7,3)$ is perpendicular to p and has the same norm as $p$. Show that $q_2 = (7,-3)$ is also perpendicular to p and has the same norm as p. Draw these points. Which one of $q_1$, $q_2$ is obtained by rotating $p$ by 90° clockwise? Which is obtained by rotating $p$ by 90° counterclockwise? Guess a formula for the 90° clockwise and counterclockwise rotation of an arbitrary point $p = (x,y)$.

(6) A formula for **R$_{90}$** is: **R$_{90}$**$(x,y) = (-y,x)$. Guess an analogous formula for **R$_{180}$**$(x,y)$ and for **R$_{270}$**$(x,y)$.

(7) Guess a formula for **H**$(x,y)$ and **V**$(x,y)$, where **H** means the horizontal flip over the $x$-axis and **V** means the vertical flip over the $y$-axis in **R**$^2$.

(8) If **R** is the rotation by 90° about the $z$-axis in $\mathbf{R}^3$, guess a formula for $\mathbf{R}(x, y, z)$.

(9) If **F** is the reflection across the $xy$-plane in $\mathbf{R}^3$, guess a formula for $\mathbf{F}(x, y, z)$. Do the same for reflections across the $xz$- and the $yz$-planes.

(10) If $\alpha$ is any angle, find a formula in terms of $\alpha$ for $\mathbf{R}_\alpha(1,0)$ and $\mathbf{R}_\alpha(0,1)$.

(11) In the sine and cosine table in the chapter, the values are rounded to 2 decimals, but these values can be determined exactly. Determine the exact values of $cos(45) = sin(45) \approx .71$ and of $sin(60) \approx .87$. *Hint: $cos(60)$ equals exactly ½.*

(12) Fill in the following table with rounded values for the sine and cosine of angles between 180 and 360. *HINT: Use the same numbers that are in the table in the chapter; namely, plus and minus 0, .26, .50, .71, .87, .97 and 1.*

| $t$ | 180 | 195 | 210 | 225 | 240 | 255 | 270 | 285 | 300 | 315 | 330 | 345 | 360 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $cos(t)$ | | | | | | | | | | | | | |
| $sin(t)$ | | | | | | | | | | | | | |

(13) Which of the following types of symmetries does the "exotic border pattern" described in this chapter have: vertical flips, horizontal flips, 180° rotations, glide-reflections.

(14) Does the exotic bounded object described in this chapter have any improper symmetries? In other words, is this object oriented?

(15) Consider the border pattern consisting of all rational numbers on the $x$-axis that can be expressed as a fraction whose denominators is a power of 2. Describe all of its translation symmetries. Does this border pattern have a "smallest translation"?

(16) The exotic bounded object described in this chapter is a countable infinite collection of points of the unit circle with the property that any point can be moved to any other point by a symmetry of the object. Is there an analogous three-dimensional object? That is, can you find a countable infinite collection of points on the sphere such that any point of the collection can be moved to any other point by a symmetry of the collection? *HINT: Because of the classification theorem, any such example must be essentially two-dimensional.*

(17) If a cube with side-length 2 is centered at the origin (0,0,0), then the locations of its eight vertices are: $(1,1,1)$, $(1,1,-1)$, $(1,-1,1)$, $(1,-1,-1)$, $(-1,1,1)$, $(-1,1,-1)$, $(-1,-1,1)$, $(-1,-1,-1)$. Draw a picture of this cube. Prove that it is a Platonic solid by verifying that all six of its faces are identical regular 4-gons. In particular, you must verify that the edges meet at right angles.

# 13. Symmetry and Matrices

In the early chapters of this book, we described rigid motions with phrases like "the 90° rotation about a point." Then we learned to describe rigid motions with formulas like **F**(x,y) = (–y,x). But for many purposes, the very best way to describe a rigid motion is with a matrix. That is what we will do in this final chapter.

## Matrix Computations

A matrix simply means a grid of real numbers.

**DEFINITION:** An *n*-by-*n* matrix means $n^2$ real numbers arranged into a square grid.

For example, here are a few 2-by-2 matrices:

$$A = \begin{pmatrix} 2 & -3 \\ 1 & 7 \end{pmatrix}, B = \begin{pmatrix} \sqrt{2} & 0 \\ -51 & 3/5 \end{pmatrix}, C = \begin{pmatrix} -20 & -3 \\ \pi & 0 \end{pmatrix}$$

and here are a few 3-by-3 matrices:

$$D = \begin{pmatrix} -2 & 5 & 9 \\ 0 & 4 & 6 \\ -2/3 & 8 & -1 \end{pmatrix}, E = \begin{pmatrix} 0 & 1 & -3 \\ 5 & -2/7 & 27 \\ 0 & \sqrt{7} & 0 \end{pmatrix}, F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

We will think of each row and each column of an *n*-by-*n* matrix as a point in $\mathbf{R}^n$. For example, the rows of the above matrix $E$ are: $(0,1,-3)$, $(5,-2/7,27)$ and $(0,\sqrt{7},0)$, while the columns of $F$ are: $(1,0,0)$, $(0,3,0)$, and $(0,0,9)$. The entries of a matrix are the numbers out of which it is built. For example, the (2,3)-entry of $D$ is 6, while the (3,2)-entry of $D$ is 8. Notice the convention of indexing the entry's row first and then its column.

K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2_13,

There are three important types of matrix computations, which we will now describe. First, we will describe how to multiply a pair of matrices. The answer is another matrix.

**HOW TO MULTIPLY A PAIR OF MATRICES:** If $A$ and $B$ are *n*-by-*n* matrices, then $A * B$ is the *n*-by-*n* matrix whose (i,j)-entry equals the dot product of the ith row of $A$ with the jth column of $B$.

**2-BY-2 EXAMPLE:**

$$\begin{pmatrix} 2 & -3 \\ 1 & 7 \end{pmatrix} * \begin{pmatrix} 3 & 4 \\ 0 & 10 \end{pmatrix} = \begin{pmatrix} (2,-3) \bullet (3,0) & (2,-3) \bullet (4,10) \\ (1,7) \bullet (3,0) & (1,7) \bullet (4,10) \end{pmatrix}$$

$$= \begin{pmatrix} 6 & -22 \\ 3 & 74 \end{pmatrix}.$$

Each of the four entries of the answer equals the dot product of the same-colored row of the first matrix with same-highlighted column of the second matrix.

**3-BY-3 EXAMPLE:**

$$\begin{pmatrix} 2 & -3 & 4 \\ 0 & 1 & 5 \\ -4 & 0 & 3 \end{pmatrix} * \begin{pmatrix} 1 & 10 & 3 \\ -2 & 4 & -1 \\ 0 & -5 & 6 \end{pmatrix} = \begin{pmatrix} 8 & -12 & 33 \\ -2 & -21 & 29 \\ -4 & -55 & 6 \end{pmatrix}.$$

As before, each of the nine entries of the answer equals the dot product of the same-colored row of the first matrix with same-highlighted column of the second matrix.

Second, we will describe how to compute the determinant of a matrix.

**THE DETERMINANT OF A 2-BY-2 MATRIX:** The <u>determinant</u> of a 2-by-2 matrix $A$, denoted as $\det(A)$, is the real number defined as:

$$det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

This is simple enough. For example,

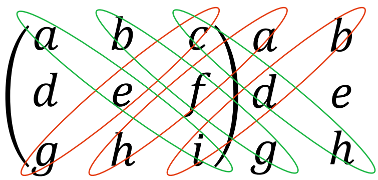
$$\det\begin{pmatrix}2 & 5\\ 7 & 3\end{pmatrix} = 2\times 3 - 5\times 7 = \ 6 - 35 = -29.$$

The determinant of a matrix is a number, which might be positive or negative or zero.

**THE DETERMINANT OF A 3-BY-3 MATRIX:** The determinant of a 3-by-3 matrix $A$, denoted as $\det(A)$, is the real number defined as:

$$det\begin{pmatrix}a & b & c\\ d & e & f\\ g & h & i\end{pmatrix} = \ aei + bfg + cdh - ceg - afh - bdi.$$

You do not need to memorize this formula because there is a simple way to visualize it. Just append copies of the first two columns of A to the right edge of A and draw red and green circles like this:




Notice that the three green circles correspond to the three positive terms in the determinant formula, while the three red circles correspond to the negative terms.

There is an analogous (but messier) formula for the determinant of a 4-by-4 matrix and a 5-by-5 matrix, and so on. In this book, you will only need to compute determinants of 2-by-2 and 3-by-3 matrices, but be aware that the word “determinant” does make sense for larger matrices.

Third, we will describe how to multiply a matrix times a point. The answer is a point:

**HOW TO MULTIPLY A MATRIX TIMES A POINT:** If $M$ is an *n*-by-*n* matrix and $p$ is a point in $\mathbf{R}^n$, then $M * p$ means the point in $\mathbf{R}^n$ whose ith coordinate equals the dot products of the ith row of $M$ with $p$.

**EXAMPLE:** If $M = \begin{pmatrix} 2 & -3 \\ 1 & 5 \end{pmatrix}$ and $p = (10, 4)$, then:

$$M * P = \big((2, -3) \bullet (10,4)\,, (1,5) \bullet (10,4)\big) = \ (8,30)$$

**EXAMPLE:** If $M = \begin{pmatrix} 1 & 2 & -3 \\ 0 & 4 & 5 \\ -2 & 0 & 1 \end{pmatrix}$ and $P = (2, 4,1)$, then:

$$M * P = \big((1,2, -3) \bullet (2, 4,1)\,, (0,4,5) \bullet (2, 4,1)\,, (-2,0,1) \bullet (2, 4,1)\big)$$

$$= \ (7,21, -3)$$

In summary, you just learned three algebraic skills: (1) multiplying a matrix times a matrix, (2) computing the determinant of a matrix, and (3) multiplying a matrix times a point. All three computations involve only adding and multiplying. In the next section, we discuss the geometric meanings of these computations, and the relationship between matrices and rigid motions.

## Representing Rigid Motions as Matrices

Matrices are great, but they are really only useful for studying the symmetries of a *bounded* object. Here is why. The familiar Center Point Theorem for two- and three-dimensional objects generalizes to objects in $\mathbf{R}^n$. The generalization says this:

**GENERAL CENTER POINT THEOREM:** Any bounded object in $\mathbf{R}^n$ has a “center point” that is fixed by each of its symmetries.

In other words, each of its symmetries leaves its center point unmoved. Any bounded object can be translated so that its center point becomes the origin $(0,0,\ldots,0)$ of $\mathbf{R}^n$. After this repositioning, all of its symmetries will be rigid motions that fix the origin (they will leave the origin unmoved). This is exactly the type of rigid motions that matrices can help us study. In this chapter, we only consider bounded objects centered at the origin. Each symmetry of such an object fixes the origin.

You previously learned to regard a rigid motion as a function and represent it using a formula. For example, $\mathbf{R_{90}}$ has the formula: $\mathbf{R_{90}}(x,y) = (-y,x)$. Now, we will redescribe a rigid motion as a matrix. Here is how.

**THE MATRIX THAT REPRESENTS A RIGID MOTION:** The matrix that represents a rigid motion, **F,** of $\mathbf{R}^n$ is the *n*-by-*n* matrix whose columns are $\mathbf{F}(1,0,\ldots,0)$, $\mathbf{F}(0,1,0,\ldots,0),\ldots,$ $\mathbf{F}(0,\ldots,0,1)$, in this order.

So the matrix representing a rigid motion, **F,** of $\mathbf{R}^2$ is the 2-by-2 matrix whose two columns are $\mathbf{F}(1,0)$ and $\mathbf{F}(0,1)$. The matrix representing a rigid motion, **F,** of $\mathbf{R}^3$ is the 3-by-3 matrix whose three columns are $\mathbf{F}(1,0,0)$, $\mathbf{F}(0,1,0)$ and $\mathbf{F}(0,0,1)$.

**EXAMPLE:** The rigid motion $\mathbf{R_{90}}$ is represented by the 2-by-2 matrix whose first column equals $\mathbf{R_{90}}(1,0) = (0,1)$ and whose second column equals $\mathbf{R_{90}}(0,1) = (-1,0)$. Thus:

$$\mathbf{R_{90}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

In what way exactly does this matrix "represent" $\mathbf{R_{90}}$? At first glance, this matrix seems to only tell us which outputs $\mathbf{R_{90}}$ associates to the inputs $(1,0)$ and $(0,1)$. Magically, this matrix also tells us the outputs $\mathbf{R_{90}}$ associate to every possible input point of $\mathbf{R}^2$. Remember how to multiply a matrix times a point? Watch:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} * (x, y) = (-y, x)$$

See what happened here? When the matrix that represents $\mathbf{R_{90}}$ was multiplied times a point of $\mathbf{R}^2$, the answer was exactly the output that $\mathbf{R_{90}}$ associates with that input point. For example:

$$\mathbf{R_{90}}(3,1) = (-1,3) \quad \leftrightarrow \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} * (3,1) = (-1,3)$$

This is exactly what always happens!

**THEOREM:** If **F** is a rigid motion of $\mathbf{R}^n$ that fixes the origin, and $M$ is the matrix that represents **F**, then for any point $p$ of $\mathbf{R}^n$, we have: $M * p$ = **F**($p$).

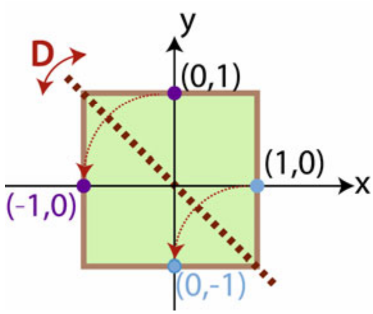
In words: "To learn the output point that **F** associates to the input point $p$, you multiply $M$ by $p$." This is a powerful theorem. For example, when *n* = 3, the columns of $M$ are defined to record where **F** moves the three points $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. It is surprising that this information is enough to determine where **F** moves ALL points of $\mathbf{R}^3$, and that this determination is achieved via something as simple as matrix-point multiplication.

Let us revisit $\mathbf{D_4}$ = the symmetry group of the square. Here are the matrices representing all eight of its members (assuming the square is centered at the origin of $\mathbf{R}^2$):

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{R_{90}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \mathbf{R_{180}} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{R_{270}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \mathbf{V} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \mathbf{D'} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Look back at the pictures of these eight symmetries in Chap. 2 and verify that these eight matrices are all correct. For example, the matrix for **D** is correct because its columns are **D**(1,0) = (0,–1) and **D**(0,1) = (–1,0) as shown in the illustration on the right.




What happens when you multiply two of these matrices? For example, what is the matrix that represents $\mathbf{R_{90}}$ times the matrix that represents **D**? Try it. You will discover the answer is the matrix that represents **H**, agreeing with the fact that $\mathbf{R_{90}}$***D** = **H**. It seems that matrix multiplication achieves composition of symmetries! Here is the general rule:

**MATRIX MULTIPLICATION ACHIEVES COMPOSITION:** If $\mathbf{F_1}$ and $\mathbf{F_2}$ are rigid motions of $\mathbf{R^n}$ that fix the origin, and $M_1$ and $M_2$ are the matrices that represent them, then $M_1 * M_2$ is the matrix that represents their composition $\mathbf{F_1}$*$\mathbf{F_2}$**.**

We now have a very effective way to translate visual questions about symmetries into algebraic questions about matrices. Do you see how it works? If your cousin Henry had never heard of $\mathbf{D_4}$, you could describe this group to him simply by showing him the eight matrices and nothing else. He could use matrix multiplication to build the Cayley table, without ever cutting out a cardboard square or visualizing a rotation or flip.

The rotations in $\mathbf{D_4}$ were all multiples of 90°, which made it easy to find the matrices representing them. To do the same for other cyclic and dihedral groups, the rotations are more complicated, so you will need this:

**THEOREM:** The matrix that represents a rotation about the origin of $\mathbf{R}^2$ by $t$ degrees is: $\begin{pmatrix} cos(t) & -sin(t) \\ sin(t) & cos(t) \end{pmatrix}$.

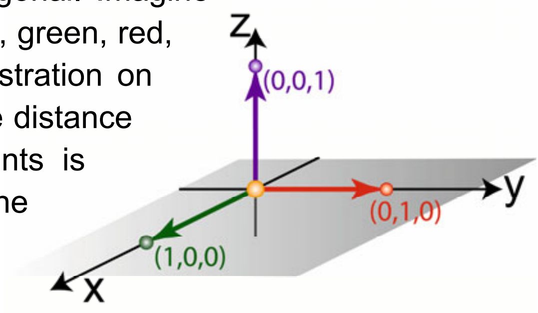
This matrix's first column is correct because cosine and sine are exactly defined so that the rotation sends $(1,0)$ to $(cos(t), sin(t))$. Think about why the second column is also correct.

## Orthogonal Matrices

Not just any matrix represents a rigid motion. The matrices that do have a special property called "orthogonal."

**DEFINITION:** A matrix is called orthogonal if the norm of each column equals 1 and the dot product of each pair of different columns equals 0.

Suppose that **F** is a rigid motion of $\mathbf{R}^3$ that fixes the origin, and $M$ is the matrix that represents **F**. Let us think about why we might expect $M$ to be orthogonal. Imagine watching **F** move the orange, green, red, and purple points in the illustration on the right. Since **F** is rigid, the distance between pairs of these points is the same before and after the motion. The orange point stays put because it is the origin (try saying "orange origin" 10 times fast). Imagine the colored arrows moving along with to aim at the new locations of the green, red and purple points. These new locations are the three columns of $M$. After the motion, the arrows will still have length 1, which is to say that the columns of $M$ have norm 1. This is simply because the distance



between the orange and each other point is unchanged by the motion. After the motion, these arrows will still be mutually perpendicular, which is to say that the dot product of each pair of different columns of $M$ equals 0. Why? If the angle between two arrows became acute, the points they aim at would have grown closer together. If the angle became obtuse, the points would have grown further apart. Can you picture this? This visual discussion is not a proof, but it should help you believe the following:

**THEOREM:** If $M$ is a matrix that represents a rigid motion that fixes the origin, then $M$ is an orthogonal matrix. Conversely, every orthogonal matrix represents a rigid motion that fixes the origin.

It turns out that orthogonal matrices have very limited possibilities for their determinants. Calculate the determinant of each of the eight matrices for $\mathbf{D_4}$. You will discover that the determinant of each rotation is 1, while the determinant of each flip is –1. The general rule is:

**THEOREM:** The determinant of any orthogonal matrix equals either 1 (if it represents a proper rigid motion) or –1 (if improper).

This theorem should perhaps be green-boxed instead of blue-boxed because it is in part a definition. It is a more precise definition of the terms “proper” and “improper” for rigid motions of the plane and space. For rigid motions of higher dimensional Euclidean spaces, it is our only definition of these terms.

What about rigid motions which do not fix the origin? Can we use determinants to distinguish whether they are proper or improper? This turns out to be easy because:

**THEOREM:** Every rigid motion of $\mathbf{R}^n$ equals a rigid motion that fixes the origin followed by a translation.

So a rigid motion is called proper or improper depending on whether its origin-fixing part has determinant equal to 1 or –1. This is a good definition. Defining "proper" using determinants is much more precise and unambiguous than our previous verbiage about right hands and clocks.

## You Finished the Book. Now What?

In this chapter, we intended only to describe how matrices are related to symmetry. We did not include any proofs because the proofs belong to a linear algebra book. We also did not carry out the important work of using matrices to prove the previously unproven theorems scattered throughout this book.

Thus, this book ends with a beginning – a more rigorous starting point from which you can revisit the topic of symmetry with more precise definitions and more complete proofs. We hope we have piqued your interest in someday reading the books and taking the classes in which this idea is fully developed. Here is a brief glimpse of some what's left to learn about symmetries and matrices.

Matrices provide a precise definition of "rigid motion" from which one can quickly prove the classifications of plane and space rigid motions found in this book. Some of the most difficult theorems in this book can then be rephrased and proved using matrices. For example, the classification of symmetry groups of solid bounded objects boils down to understanding the possible finite subgroups of the group of all orthogonal 3-by-3 matrices.

Things get really interesting when you move into higher dimensional Euclidean spaces. Consider this question:

Classify the possible symmetry groups of bounded objects in $\mathbf{R}^n$.

You already know the answer for $n = 2$ and $n = 3$. What about general $n$? This question turned out to be very difficult and very important. It motivated some of the most significant mathematics of the past century. First, this question is intertwined with the problem of classifying all possible finite groups, which lead to one of the most celebrated achievements of modern mathematics, aptly called "The Enormous Theorem." Second, this question is related to the classification of compact Lie groups, which is a beautiful piece of mathematics upon which much of modern physics depends.

Alternately, there is a very modern method for proving classification theorems without matrices. This modern approach, championed by John Conway and others, uses orbifold geometry to classify the ways in which border patterns, wallpaper patterns, and bounded solid objects can be symmetric. A general-audience introduction of this program is found in a beautiful book titled *The Symmetry of Things,* by Conway, Burgiel and Goodman-Strauss.

## Exercises

(1) Consider: $A = \begin{pmatrix} 2 & -3 \\ 1 & 7 \end{pmatrix}$, $B = \begin{pmatrix} 3 & 0 \\ -5 & 1/2 \end{pmatrix}$, $C = \begin{pmatrix} -10 & -3 \\ 4 & 0 \end{pmatrix}$.

(a) Compute the determinant of each matrix.

(b) Compute $A * B$ and $B * A$. Are they equal?

(c) Verify that $det(A * C) = det(A) * det(C)$.

(d) Compute $A * (2,3)$ and $A * (-2,7)$.

(e) Verify that $A$ is not orthogonal.

(f) The function from $\mathbf{R}^2$ to $\mathbf{R}^2$ determined by $A$ (sending $P$ to $A * P$) is not a rigid motion because $A$ is not orthogonal. Show this directly by finding a pair of points whose distance is not preserved.

(2) Consider: $A = \begin{pmatrix} -2 & 5 & 9 \\ 0 & 4 & 6 \\ -1 & 2 & -1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 & -3 \\ 5 & -2 & 10 \\ 0 & 1 & 0 \end{pmatrix}$.

(a) Compute the determinants of $A$ and $B$.

(b) Compute $A * B$ and $B * A$. Are they equal?

(c) Compute $A * (1,2,3)$ and $A * (0,4,-2)$.

(d) The function from $\mathbf{R}^3$ to $\mathbf{R}^3$ determined by $A$ (sending $P$ to $A * P$) is not a rigid motion because $A$ is not orthogonal. Show this directly by finding a pair of points whose distance is not preserved.

(3) Write the matrix that represents each of the six symmetries in $\mathbf{D_3}$.

(4) Determine the matrix that represents each of the following rigid motion of $\mathbf{R}^3$:

(a) The 90° rotation about the $z$-axis or the $x$-axis.

(b) A reflection across the $yz$-plane or the $xz$-plane.

(5) In this problem, we will revisit the central inversion rigid motion defined at the end of Chap. 7.

(a) Find the 3-by-3 matrix, $M_1$, that represents $\mathbf{F_1}$ = the reflection across the $xy$-plane.

(b) Find the 3-by-3 matrix, $M_2$, that represents $\mathbf{F_2}$ = the 180° rotation about the $z$-axis.

(c) Verify that $\mathbf{F_1}$ and $\mathbf{F_2}$ commute by checking that $M_1 * M_2 = M_2 * M_1$. Can you picture this using the illustration found in Chap. 7?

(d) The composition $\mathbf{F} = \mathbf{F_1} * \mathbf{F_2}$ is called <u>central inversion</u>. Find the matrix $M = M_1 * M_2$ that represents **F**. Verify that **F** has the formula $\mathbf{F}(x, y, z) = (-x, -y, -z)$.

(e) Verify that **F** is improper by showing $det(M) = -1$.

(f) Verify that **F** commutes with every rigid motion of $\mathbf{R}^3$ that fixes the origin. Do this by checking that $A * M = M * A$ for every 3-by-3 matrix $A$.

(g) Explain why F is a symmetry of the cube and the dodecahedron but not of the tetrahedron.

(6) Describe in words the rigid motion of $\mathbf{R}^3$ represented by:

$$\begin{pmatrix} \cos(t) & 0 & -\sin(t) \\ 0 & 1 & 0 \\ \sin(t) & 0 & \cos(t) \end{pmatrix}$$

# Index



K. Tapp, *Symmetry: A Mathematical Exploration*, DOI 10.1007/978-1-4614-0299-2, 