LETTER FROM THE EDITOR

The cover, entitled *Chicken Feet*, does a good job of introducing a few of the themes in this "summer fun" issue. The cover was designed by David Reimann and displays an array of chicken foot quilting patterns. In the first article of this issue, David McCune and Lori McCune use Markov chains to analyze the *Count Your Chickens!* board game. The game is billed as a "cooperative board game" and its play is a bit unusual. In the second article, Kathryn Haymaker and Beth Malmskog examine a round robin quilting circle in which each person starts their own quilt by making a square (such as a chicken foot). They use row complete Latin squares to create a scheme so that each person passes a quilt to every other person in the circle under certain constraints.

In the next article, Jathan Austin, Brian Kronenthal, and Susanna Molitoris Miller consider a more advanced board game: *Catan*. *Catan* (formerly *The Settlers of Catan*) is a property-building and trading board game. Players build a new board every time they play by arranging tiles, number tokens, and port markers. The authors use combinatorics, probability, and abstract algebra to answer questions about the number of possible boards.

Keeping the summer fun going, Jordan Schettler explores Wendy Carlos's xenharmonic musical scales, explaining how Carlos's scales arise naturally from the theory of continued fractions. He shows that it is possible to construct an infinite family of similarly "nice" scales using the same approach.

Now, back to chickens, or in this case, the game of chicken. Kimmo Eriksson and Jonas Eliasson introduce the Chicken Braess paradox. The Braess paradox is the counterintuitive fact that the addition of a link to a network may degrade performance. For a road network, Eriksson and Eliasson consider the addition of a single-lane, but two-way street. Which drivers get to use the road is an example of the game of chicken, resulting in a new twist to the Braess paradox. Their article concludes with a number of exercises that could lead to some interesting student work.

In the final article of the issue, Zili Wu touches on one of my favorite Calculus II topics: convergence tests for series. Wu considers two convergence theorems that extend the ratio test for p-series.

Sprinkled throughout the issue is a proof without words of the volume of a regular tetrahedron by Adrian Chun Pong Chu and a Math Bite about hobbits and a birthday-like problem by Stanley R. Huddy. Brendan Sullivan has once again provided a crossword puzzle with a MathFest theme to help you prepare for the Mathematical Association of America's summer meeting. David Nacin provides a TRIBUS puzzle as a fun way to exercise your brain over the break. More strenuous exercises appear in the Problems section. To conclude the issue, the Reviews section gives some suggestions for summer reading, including a book by Frank Zagare on game theory and diplomacy.

Michael A. Jones, Editor

ARTICLES

Counting Your Chickens With Markov Chains

DAVID MCCUNE
William Jewell College
Liberty, MO 64068
mccuned@jewell.edu
LORI MCCUNE
Missouri Western State University
St. Joseph, MO 64507
lmccune@missouriwestern.edu

There are many upsides to having children: the experience of joy (*joy* as distinct from *happiness*), the avalanche of cuteness, and discovering the exact depth of your well of patience. There are also some downsides: sleepless nights, a complete lack of personal space when using the bathroom, and the Sisyphean futility of housecleaning. A downside that we did not anticipate is that many games for children are boring. The reason is that children's games are often games of pure chance. There are two ways to make a game of pure chance interesting: either introduce gambling or think about the mathematics behind the game. Our children don't have any money, so in order to make their games interesting we have opted for the latter option. In this article we use Markov chains to analyze one of our daughter's favorite games of chance, a cute board game called *Count Your Chickens!*. We calculate the probability of winning the game for the original game board and for many variants of the board. Our focus on win probability distinguishes our work from most of the literature that uses Markov chains to analyze games, where the results primarily concern expected game length.

Count Your Chickens!

Markov chains have been used to study many board games such as Risk [5], Chutes and Ladders ([1], [2]), Hi Ho Cherry-O [4], and others. We use Markov chains to study a game aimed at preschoolers, *Count Your Chickens!*, produced by Peaceable Kingdom. In this cooperative game, gameplay begins with forty chicks spread throughout the board and a mama chicken at the start square. The end square is the mama chicken's coop, and the mama chicken moves progressively closer to this square as the game unfolds. The object of the game is to have all 40 chicks in the chicken coop when the mama chicken arrives there. The game has 41 squares numbered in order, with 30 squares containing one of five characters: a cow, a sheep, a pig, a dog, or a tractor. Five of these squares are colored blue. The final square, the chicken coop, contains all 5 characters. Players take turns spinning a spinner with 6 segments of equal area, five of which match the characters from the board and a sixth segment containing a fox. When a player spins the spinner, one of two things can happen. If the player spins one of the five characters drawn on the board, she moves the mama chicken token to the next iteration on the path to the coop of the character spun. The player counts the number of squares moved, and this number of chicks is placed in the chicken coop. If

Math. Mag. **92** (2019) 163–172. doi:10.1080/0025570X.2019.1561097 © Mathematical Association of America
MSC: Primary 91A60; Secondary 60J10
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.
Supplementary material for this article can be accessed on the publisher's website.

the square landed on is blue, the player adds one additional chick to the coop. There are only a total of 40 chicks on the board though, so if the mama chicken moves two squares but there are already 39 chicks in the coop, only one chick gets added. If a fox is spun, the player must remove one chick from the coop and place it back on the board. If there are no chicks in the coop then spinning a fox has no effect.

For example, suppose the mama chicken is at the start square. If we spin a sheep, then the mama chicken would move to square 2 (the first sheep) and we place two chicks in the coop (see figure 1). If we next spin a tractor then the mama chicken would move two squares to square 4 and we would place three more chicks in the coop because square 4 is blue. If our first spin is a fox then nothing happens because at the start of the game there are no chicks in the coop. Gameplay continues in this manner until the mama chicken reaches the coop. The players win if all 40 chicks are in the coop when the mama chicken arrives there; otherwise the players lose.

Playing this game has been a good way for our daughter to practice her counting and addition and subtraction. Given her propensity for cheating, it's not clear how much these lessons sink in. To make the game interesting for us, we decided to answer the following questions.

1. What is the probability of winning a game of *Count Your Chickens!*? That is, what is the probability of landing on square 40 with 40 chicks in the coop?
2. What is the expected number of chicks in the coop at the end of the game?
3. What effect do the blue squares have on the probability of winning the game? For example, the original game has five particular squares colored blue; are there other choices of five blue squares that increase the probability of winning the game?

All of these questions can be answered using a classic probabilistic tool called a *Markov chain*. Each question requires its own particular chain or class of chains.

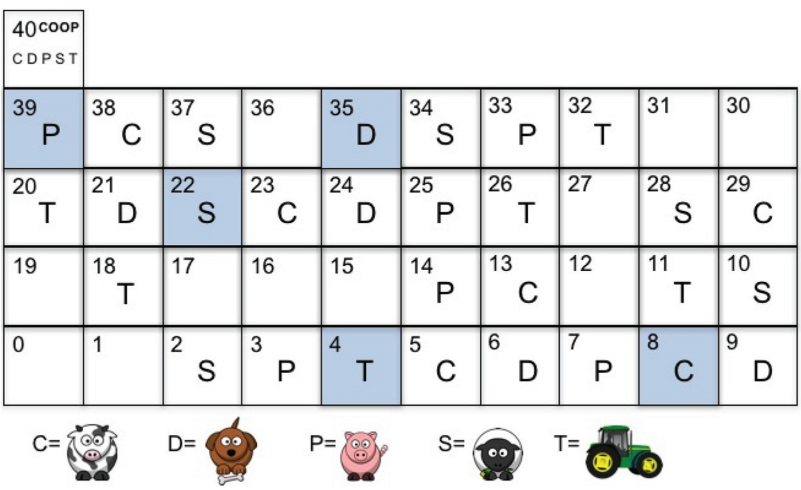


Figure 1 *Count Your Chickens!* game board with blue squares shaded; characters from https://openclipart.org.

Building a Transition Matrix to Analyze the Game

The transition matrix P for a Markov chain is constructed so that entry p_{ij} of the matrix is the probability of transitioning from state i to state j. For our purposes, a state in *Count Your Chickens!* consists of an ordered pair (a, b) where a is the mama chicken's position on the board and b is the number of chicks in the coop. When we are interested only in win probabilities then, to minimize the number of states and cut down on computation time, we include a "loss" state and a "win" state. The idea is that while (39, 35), for example, is a possible state of the game, it is impossible to win from such a configuration. All such ordered pairs from which it is impossible to win are grouped together as a loss state. The win state is (40, 40)—the mama chicken must reach the last square of the board with all of the chicks in the coop.

We illustrate the construction of such a chain with a smaller board in the following example.

Example 1. In the smaller game, there are two characters on the board—a sheep and a cow. There are 8 squares on the board after the start square and 8 chicks to return to the coop. The spinner has three equal-sized segments illustrated with a cow, a sheep, and a fox. We number only the squares with a character, since it is not possible to land on a blank square.

Figure 2 Smaller game board for Example 1.

As described above, a state of the game is an ordered pair (i, j) where i is the position of the mama chicken and j is the number of chicks in the coop. Notice that some such pairs will be identified with the loss state because, if j is small enough relative to the mama chicken's position, it will be impossible for all 8 chicks to be in the coop at the end of the game. The state (2, 0) is such an example: if the mama chicken is on square 2 and there are no chicks in the coop, the game will end with a maximum of 7 chicks in the coop. Let b_i^R be the number of blue squares after square i on the path and b_i^L the number of blue squares that come before square i, including square i. Note $b_i^R + b_i^L$ is the total number of blue squares in the game. Then for square i, we get the states (i, j) where $\max(i - b_i^R, 0) \leq j \leq \min(i + b_i^L, 8)$. For example, for square 3, $b_3^R = 1$ and $b_3^L = 1$ since there is one blue square to the right of square 3 and one blue square to the left of square 3. Thus for square 3 we get the possible states (3, 2), (3, 3), (3, 4). Note that it is not possible to have more than 4 chicks in the coop if you are on square 3, and if you have 0 or 1 chicks in the coop when you are on square 3 it is impossible to win the game. Hence we get states

$$\{(0,0), (2,1), (2,2), (2,3), (3,2), (3,3), (3,4),$$
$$(5,5), (5,6), (5,7), (7,7), (7,8), \text{ win}, \text{ loss}\}$$

for this board. For the purposes of building a transition matrix we think of (0, 0) as state 1, (2, 1) as state 2, (2, 2) as state 3, ..., and "loss" as state 14.

Entry p_{ij} in the transition matrix for the game is the probability of moving from state i to state j. For example, in row one of the transition matrix we start in position (0, 0) on the gameboard. We have a 1/3 probability of rolling a fox; this spin keeps us in position (0, 0), so $p_{1,1} = 1/3$. We could also spin a sheep, putting us in position (2, 3), or a cow, putting us in position (3, 3). This gives $p_{1,4} = 1/3$ and $p_{1,6} = 1/3$. All other entries in row one are 0. If we start in state (2, 1) then spinning the sheep would put us in position (7, 6), or the loss state, because if we have only six chicks in the coop when we reach square seven, it is impossible to win. Spinning the fox also would put us in the loss state, so entry $p_{2,14} = 2/3$. Spinning the cow puts us in state $(3, 2)$, so $p_{2,5} = 1/3$. All other entries in row two are 0. Continuing in this way, we find the transition matrix for this game is

$$P = \begin{bmatrix} \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

In the original game, for each of the 30 numbers with a character on its square in the original gameboard we get states (i, j) for $\max(i - b_i^R, 0) \leq j \leq \min(i + b_i^L, 40)$. The matrix is too large to display, but in form it resembles that of the small example above with multiples of $1/6$ as the nonzero entries because the spinner has six equally likely outcomes.

Win Probability and Expected Number of Chicks

The matrix in equation (1) models the game in Example 1 as an absorbing Markov chain—the loss and win states are both absorbing (each of these states transitions to itself with probability 1), and it is possible to reach one of these two states from any other state. The transition matrix for any absorbing Markov chain can be written in the canonical form

$$\begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

where I is an identity matrix. Notice that the transition matrix for our example (equation 1) is in canonical form because we listed the two absorbing states as the final two states. In general, any transition matrix for an absorbing chain can be written in canonical form simply by listing the absorbing states at the end of a list of states. With the transition matrix in canonical form, we define the *fundamental matrix* N by $N = (I - Q)^{-1}$. Let $B = N \times R$. If the game is in the ith non-absorbing state then the probability that the game terminates by reaching the jth absorbing state is b_{ij}. For details and proofs, see [**3**]. In our small example,

$$N = \begin{bmatrix} \frac{3}{2} & \frac{1}{18} & \frac{1}{6} & \frac{1}{2} & \frac{2}{9} & \frac{11}{18} & \frac{1}{6} & \frac{4}{27} & \frac{2}{9} & \frac{1}{18} & \frac{67}{162} & \frac{17}{54} \\ 0 & 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{9} & 0 & 0 & \frac{1}{27} & 0 \\ 0 & \frac{1}{3} & 1 & 0 & \frac{2}{9} & \frac{1}{3} & 0 & \frac{1}{9} & \frac{1}{9} & 0 & \frac{40}{81} & \frac{1}{27} \\ 0 & \frac{1}{9} & \frac{1}{3} & 1 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{2}{27} & \frac{1}{9} & \frac{1}{9} & \frac{31}{81} & \frac{14}{27} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{9} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 1 & 0 & \frac{2}{9} & \frac{1}{3} & 0 & \frac{4}{9} & \frac{1}{9} \\ 0 & 0 & 0 & 0 & \frac{1}{9} & \frac{1}{3} & 1 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} & \frac{26}{81} & \frac{14}{27} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 1 & 0 & \frac{2}{9} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{9} & \frac{1}{3} & 1 & \frac{5}{27} & \frac{4}{9} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 1 \end{bmatrix} \text{ and } R = \begin{bmatrix} 0 & 0 \\ 0 & \frac{2}{3} \\ 0 & 0 \\ 0 & 0 \\ 0 & \frac{2}{3} \\ 0 & 0 \\ 0 & 0 \\ \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 \\ \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & 0 \end{bmatrix}.$$

Multiplying N times R gives a 12×2 matrix whose $(i, 1)$ entry gives the probability of reaching the win state starting from state i and the $(i, 2)$ entry gives the probability of reaching the loss state starting from state i. We get

$$B = \begin{bmatrix} 0.6276 & 0.3724 \\ 0.0617 & 0.9383 \\ 0.4280 & 0.5720 \\ 0.6996 & 0.3004 \\ 0.1852 & 0.8148 \\ 0.5556 & 0.4444 \\ 0.7819 & 0.2181 \\ 0.5556 & 0.4444 \\ 0.8148 & 0.1852 \\ 0.9012 & 0.0988 \\ 0.6667 & 0.3333 \\ 0.8889 & 0.1111 \end{bmatrix}.$$

From the first row of B, this game has a win probability of 0.6276 and a loss probability of 0.3724 when a player begins from the start square.

As a side-note, we mention that one of the nice aspects of using a Markov chain for these calculations is that a player can update her win probability as she moves through the game. For example, suppose the player starts by spinning a sheep and then spins a fox. This state of the game corresponds to the state (2, 1), which corresponds to the second row of B. Therefore in such a situation the player would know that her win probability is now only 6.17%.

The computations for the general game are quite similar except that the matrices Q and R are much larger. Q is 163×163 and R is 163×2 and therefore B is 163×2. It is impractical to construct such large matrices by hand, so we used a C++ program with the Eigen linear algebra package to build the matrices and compute the win probability. (The program is available as a supplement to this article.) The matrix Q in the actual game looks similar to the matrix Q in our example: the matrix is mostly zeroes punctuated sparsely with small diagonals of entries of value $1/6$, instead of $1/3$. The top left entry of B in this case is 0.6410, and therefore we know that the probability of winning a game of *Count Your Chickens!* is 64.10%.

If we wish to know the expected number of chickens in the coop at the end of the game then we can still use a Markov chain where states are ordered pairs as above, but we cannot collapse all of the losing states to a single loss state. This trick saved us some computational time and complexity when finding the win probability; we can't save such time for this question. Doing so would cause us to lose data about the precise number of chicks in the coop at a given stage of the game. Therefore the state set of this Markov chain is $\{(i, j) : 0 \le i \le 40, 0 \le j \le \min\{i + b_i^L, 40\}\} - \{(40, 0)\}$. The reason we throw out (40, 0) is that it is impossible to end the game with zero chicks in the coop. With this set of states, the number of rows in Q rises from 163 to 668 and the number of absorbing states rises from 2 to 40. The matrices Q and R are now even more impractical to build by hand, so again we turned to a C++ program. The first row of B gives the probabilities for arriving at any of the absorbing states. To calculate the expected number of chicks in the coop at the end of the game we scale each of these probabilities by the number of chicks corresponding to that state. Doing so gives us a value of 39.22 chicks.

We summarize these findings in the following theorem.

Theorem. *The probability of winning a game of* Count Your Chickens! *is 64.10% and the expected number of chicks in the coop at the end of the game is 39.22 chicks.*

Changing the Blue Squares

In this section we examine how the choice of blue squares affects the win probability of the game. In particular, we would like to investigate the following two hypotheses that seem intuitively clear:

1. Given a choice between more blue squares and fewer blue squares, it seems clear that in order to maximize win probability we should choose more blue squares.
2. If we fix the number of blue squares then it seems clear that the way to maximize (respectively minimize) the win probability is to color blue the squares that have the highest (respectively lowest) probability of being landed on during the course of a game.

To streamline our language below, for a given square we say that its *landing probability* is the probability of landing on that square during the game. So hypothesis (2) says that in order to maximize (respectively minimize) the win probability for a fixed number k of blue squares, we should color blue the k squares of largest (respectively smallest) landing probability.

Example. In our small example pictured in Figure 2, suppose that only square 2 was colored blue. The state set for the Markov chain is now

$$\{(0, 0), (2, 2), (2, 3), (3, 3), (3, 4), (5, 5), (5, 6), (7, 7), \text{ win}, \text{ loss}\}.$$

Note that we omit states (2, 1) and (3, 2) because these states are no longer viable (that is, they have been identified with the loss state). The transition matrix is now smaller because these two states are gone. Once we have a transition matrix we can compute a win probability using the matrix $B = N \times R$ as before. In Table 1 we record the win probability for the given sequence of blue squares in the smaller game board with 8 chicks.

TABLE 1: Win probabilities for choices of blue squares.

Blue squares	Win probability	Blue squares	Win probability
2	0.5288	3, 5	0.6684
3	0.6029	3, 7	0.7579
5	0.4918	5, 7	0.7106
7	0.6214	2, 3, 5	0.7761
2, 3	0.7222	2, 5, 7	0.7750
2, 5	0.6276	2, 3, 7	0.8100
2, 7	0.6975	3, 5, 7	0.8069

Notice that in this example our first hypothesis is true: more blue squares is better than fewer blue squares for the win probability. What about our second hypothesis: if we fix a number of blue squares k, is the win probability maximized by coloring blue the k squares of largest landing probability? To answer this question we first need to know the landing probability of each square. These probabilities can be calculated using small Markov chains in which we do not keep track of the number of chicks. For example, to calculate the landing probability of square 3 we create a Markov chain with state set $\{0, 2, 3, \text{went\_past\_3}\}$ where state 3 and state went_past_3 are made to be absorbing; remember that we cannot land on square 1. Markov chains make computing these probabilities much easier than using the direct method of finding all possible ways of landing on square 3 and adding up the corresponding probabilities. To compute the landing probability of square 3 directly, note that to reach square 3 we can either first pass through the sheep on square 2 or not pass through that square. To travel to square 3 without first landing on square 2, we'd need to spin some number of foxes and then the cow. So the probability of landing on square 3 in this way is given by the geometric series $\sum_{i=0}^{\infty} \frac{1}{3}\left(\frac{1}{3}\right)^i = \frac{1/3}{1-1/3} = \frac{1}{2}$. To find the probability of landing on square 3 given that we first land on square 2, we'd multiply the geometric series $\left(\sum_{i=0}^{\infty} \frac{1}{3}\left(\frac{1}{3}\right)^i\right)\left(\sum_{i=0}^{\infty} \frac{1}{3}\left(\frac{1}{3}\right)^i\right) = \frac{1}{4}$. Putting this together, we get that the probability of landing on square 3 is $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.

This landing probability was a bit cumbersome to compute this way and the computations become more tedious the farther the square of interest is from the start square. The Markov chain described in the previous paragraph simplifies the computation of this probability. Using the states described, we obtain matrices

$$P = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

The top left entry of B tells us that the landing probability of square 3 is 0.75.

Using four such chains we obtain the landing probabilities for squares 2, 3, 5, and 7, as shown in Table 2. Our intuition appears to be correct for this second hypothesis also: if we fix a $k \in \{1, 2, 3\}$ then we maximize the win probability by coloring blue the k squares of largest landing probability; likewise, our conjecture about how to minimize the win probability holds for this example, too.

TABLE 2: Landing probabilities for squares in the small example.

Square	Landing probability
2	0.5000
3	0.7500
5	0.3750
7	0.8125

TABLE 3: Choices of one to six blue squares that give either a minimal or maximal win probability.

Blue squares	Min win probability	Blue squares	Max win probability
20	0.2312	35	0.3607
20, 37	0.2711	10, 35	0.4731
9, 20, 37	0.3199	10, 11, 35	0.5788
9, 18, 20, 37	0.3734	6, 10, 11, 35	0.6640
8, 9, 18, 20, 37	0.4260	6, 10, 11, 22, 35	0.7292
8, 9, 18, 20, 32, 37	0.4771	6, 10, 11, 14, 22, 35	0.7850

To investigate how the choice of blue squares affects the win probability in the original game, we again used a C++ program to calculate the probabilities; the program is available as a supplement to this article. For a fixed $k \in \{1, 2, 3, 4, 5, 6\}$, the program goes through all possible choices of k blue squares. Since there are 29 viable squares for blue squares (we didn't allow the coop to be colored blue), once we made our choice of k we had the program run through all possible $\binom{29}{k}$ choices of blue squares and determine a choice that gives the minimal win probability for k blue squares and a choice that gives the maximal win probability for k blue squares; see Table 3.

Notice that our first hypothesis breaks down for the original game. Coloring only the squares 10 and 35 blue gives a higher win probability than coloring the squares 8, 9, 18, 20, and 37 blue. More blue squares doesn't necessarily translate to a higher win probability. The original game board is large enough that we can have some really awful squares (like square 20) and some really great squares (like square 35), so having just a couple of high landing probability blue squares can be better than having a lot of low landing probability blue squares.

To investigate our second hypothesis we need to know the landing probabilities of squares in the original game. These probabilities can be calculated using the same logic that we used in the small example. As with our other calculations, we used a C++ program to obtain these probabilities. The six squares with lowest landing probabilities and the six squares with highest landing probabilities in the actual game are displayed in Table 4.

Our hypothesis is true for maximal win probabilities when using five or fewer blue squares. However, our intuition breaks down with 6 blue squares. The win probability for the six squares of highest landing probability is 0.7836, slightly less than the maximal win probability shown in the table.

TABLE 4: The six squares of highest and lowest landing probability in the original game.

Square	Landing probability	Square	Landing probability
20	0.0405	25	0.3829
37	0.1421	22	0.4140
9	0.1752	6	0.4147
2	0.2000	11	0.4219
32	0.2020	10	0.4250
18	0.2026	35	0.4742

To see why this may occur, consider the case of 8 blue squares. The 8 squares of highest individual landing probability are 26, 21, 25, 22, 6, 11, 10, and 35, and the corresponding win probability is 0.8633. If we color the squares 6, 10, 11, 14, 21, 22, 25, and 35 blue then the win probability is 0.8690. If we color squares 5, 6, 10, 11, 14, 21, 22, and 25 blue then the win probability is 0.8641. What seems to be occurring is that individual landing probability needs to be balanced against "bunching" of blue squares. To make this point clear, suppose that squares 2, 3, and 4 were the squares of highest landing probability. If we landed on square 2 then the probability of skipping the next two blue squares on the next spin is 0.5. If blue squares are too closely grouped together then it becomes unlikely that we will land on all or even most of them; we can skip past many of them in a single spin.

Our hypothesis breaks down more quickly for minimal win probability. For $k = 1, 2$, and 3 blue squares, coloring the k squares of smallest landing probability achieves the minimal win probability. However, beginning with $k = 4$, coloring the k squares of smallest landing probability does not result in the minimal win probability.

Further Exploration and Conclusion

There are many other questions an interested reader could investigate. For example, what effect does adding additional animals to the board (i.e., increasing the size of the spinner) have? What if the fox removes more than one chick from the coop? What effect does changing the size of each character's space in the spinner do? For example, what if the fox covered 1/3 of the spinner and the remaining five animals each took up 2/15 of the spinner? What if some of the blue squares were colored green and had the effect of moving an additional two chicks into the coop? We invite you to create your own questions that can be analyzed using Markov chains and to modify the supplemental C++ code to answer the questions.

To conclude, we feel we should mention that none of the above calculations actually apply to our personal copy of *Count Your Chickens!*. Our daughter has lost a couple of chicks, so we only need 38 chicks in the coop at the end to win the game. That is, our winning state is (40, 38) and, using the same Markov chain computations as above, we can say that the win probability for our copy of the game is 81.77%.

When we reported all of our findings to our daughter, she seemed completely indifferent. We think this is because she has found a strategy with a win probability of 100%: at any point during the game she moves as many chicks into the coop as she wants, and she'll flip the spinner at you if you protest.

REFERENCES

[1] Althoen, S. C., King, L., Schilling, K. (1993). How long is a game of Snakes and Ladders? *Math. Gazette* 77: 71–76. doi.org/10.2307/3619261

[2] Cheteyan, L. A., Hengeveld, S., Jones, M. A. (2011). Chutes and Ladders for the impatient. *College Math. J.* 42: 2–8. doi.org/10.4169/college.math.j.42.1.002

[3] Grinstead, C. M., Snell, J. L. (1997). *Introduction to Probability*. 2nd ed. Providence, RI: American Mathematical Society. http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf

[4] Johnson, R.W. (2003). Using games to teach Markov chains. *PRIMUS: Probl., Resourc. Issues Math. Undergraduate Stud.* XIII: 337–348. doi.org/10.1080/10511970308984067

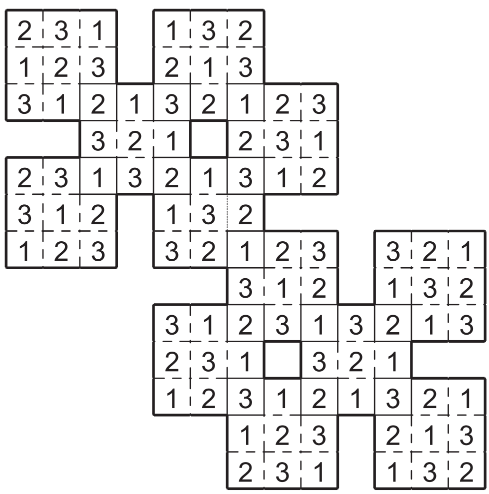
[5] Pierce, P., Wooster, R. (2015). Conquer the world with Markov chains. *Math. Horizons* 22 (4): 18–21. doi.org/10.4169/mathhorizons.22.4.18

Summary. We use Markov chains to analyze the cooperative children's game, *Count Your Chickens!*. Similar to *Chutes and Ladders* or *Monopoly*, this game involves movement of a token on a game board in which the movement is determined by pure chance. We use Markov chains to determine the win probability for the game and the win probability for variants of the game in which we change the structure of the underlying game board.

DAVID MCCUNE (MR Author ID: 792485) received his PhD in mathematics from the University of Nebraska-Lincoln in 2011. He is currently an assistant professor of mathematics at William Jewell College. He looks forward to when his children have money of their own; games of pure chance should be more interesting then.

LORI MCCUNE (MR Author ID: 1050945) received her PhD in mathematics from the University of Nebraska-Lincoln in 2011. She is currently an assistant professor of mathematics at Missouri Western State University. When she's not doing mathematics, she enjoys playing board games with her husband and two small children.

TRIBUS Solution



What (Quilting) Circles Can Be Squared?

BETH MALMSKOG
Colorado College
Colorado Springs, CO 80903
beth.malmskog@gmail.com

KATHRYN HAYMAKER
Villanova University
Villanova, PA 19085
kathryn.haymaker@villanova.edu

Judy's question. On a lucky morning in 2012, I received the following enticing email from a friend.

Dear Beth,

I have a math problem for you! I am organizing a round robin quilt exchange. That means each person in a group of quilters makes a small quilt block, and then the quilters pass the blocks to each other over a period of time until everyone in the group has added a border to everyone else's quilt. We hang on to each quilt for about a month, and each person needs to have a quilt to work on every month.

The last time we did this activity we basically passed our projects in a big circle. We each passed to the same person each time until we got our own quilt back. It was kind of limiting, since part of the fun of doing these trades is getting to meet the other people in the group. I'd like to arrange the trades so no one ever gets their quilt from the same person, but I can't quite figure it out. I tried it with 5 people in a group and, while everyone gets the quilt to sew on, the quilts are passed twice to the same person.

I think we will pretty much always have 5 people in a group, but it is possible that we would have groups with 4 or 6 people. Is there something that will work for that, too? Thanks,

Judy

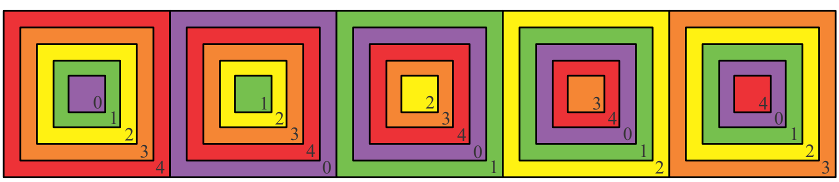


Figure 1 Five quilts created by passing in a cycle. This method was unsatisfactory because each quilter always passed to the same neighbor, which can be seen in the repeated color transitions in the quilts. For instance, purple (color 0) is immediately followed by green (color 1) in four of the quilts.

Who could resist? This email marked the start of a journey through experimental design, combinatorics, graph theory, and group theory with my mathematical traveling companion Katie Haymaker, and finally led us to a world of fascinating unsolved problems. The first step on the journey: a cup of coffee. Next step: investigating latin squares.

Math. Mag. **92** (2019) 173–186. doi:10.1080/0025570X.2019.1588033 © Mathematical Association of America
MSC: Primary 05B15
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

The Science of Squares

Let S be a set of n distinct elements—for example, S may consist of n people, or the integers 0 through $n-1$. An n by n array filled with elements from S such that each element appears once in each row and once in each column of the array is called a *latin square* of order n.

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

Figure 2 The latin square of order 5 associated to the quilts in Figure 1. Each row represents the path of a different quilt through the quilters: the sequence of colors in a row of the square corresponds to the sequence of colors in the quilt, starting in the center and proceeding outward.

A round robin quilt exchange can be modeled with a latin square. Think of each row as the path of a quilt through the quilters, labeled 0 through $n-1$. Let i and j be elements of S. The condition that no quilt undergoes the same pass twice creates the condition that i should be followed directly by j when reading across the rows (i.e. the sequence (i, j) appears) at most once in the square. There are $n(n-1)$ different sequences (i, j), and $n(n-1)$ quilt passes must take place, so in fact each sequence (i, j) must appear exactly once in the array.

An $n \times n$ latin square with this property is referred to as a *row complete latin square* of order n, abbreviated RCLS(n). Note that the square in Figure 2 is not row complete, since the purple (color 0) is directly followed by green (color 1) in many rows. Row completeness was first studied by Williams in 1948, when developing designs for experiments in which treatments might have residual effects, for example, the milk production of dairy cows on various feeds [**20**]. Residual effects are also a concern in taste-testing experiments [**5**].

Williams himself devised a construction of order n row complete latin squares for even n as follows. Say that the items in the set S are the numbers 0 through $n-1$. The *successive difference* corresponding to the sequence (i, j) is $j - i \mod n$. Then each row of an $n \times n$ latin square has a sequence of $n-1$ successive differences. For example if $n = 6$, and a row of the square consists of the sequence $(0, 1, 5, 2, 4, 3)$, the corresponding sequence of successive differences is $(1, 4, 3, 2, 5)$, which is the same as $(1, -2, 3, -4, 5)$ modulo 6.

We call a latin square *rotational* if each row has the same sequence of successive differences modulo n (such designs are sometimes also called cyclic [**5**]; however this has a different meaning in the context of latin squares). In terms of the quilt exchange, this means that there is essentially a single passing pattern that all quilts follow. Williams noticed that the sequence of successive differences $(1, -2, 3, -4, 5, \ldots, -(n-2), (n-1))$ leads to a rotational RCLS(n) for even n.

Following Williams pattern for $n = 6$, the successive differences are $(1, 4, 3, 2, 5)$ as mentioned above. Starting with 0 in the upper left entry of the square, and successively adding 1 for the first entry of the next row, we get the latin square in Table 1.

0	1	5	2	4	3
1	2	0	3	5	4
2	3	1	4	0	5
3	4	2	5	1	0
4	5	3	0	2	1
5	0	4	1	3	2

TABLE 1: An example of Williams' construction of an RCLS(6).

We call any latin square that has the sequence $(0, 1, \ldots, n-1)$ in the first column and with successive differences $(1, -2, 3, -4, 5, \ldots, -(n-2), (n-1))$ along each row a *Williams latin square*.

For an order n latin square to be row complete, the successive differences must all be distinct mod n. We can see that Williams' squares will always be latin, but will not be row complete for odd n. This is because when n is odd there are repeats among the list of successive differences $(1, -2, 3, -4, 5, \ldots, (n-2), -(n-1))$. For example, $1 \equiv -(n-1) \mod n$ for odd n.

On the impossibility of squaring some quilting circles

Can we create an RCLS(n) for any n? A quick experiment reveals that the answer is no. Consider the possibilities for a 3×3 array, denoted by Q. Let $Q_{i,j}$ denote the entry in the ith row and jth column of the array. One may rename the entries and reorder the rows so that entries in the first row and column follow the sequence (0, 1, 2).

$Q =$

0	1	2
1	?	
2		

The first entry up for debate is $Q_{2,2}$. The entry $Q_{2,2}$ must be a 0 or a 2 since there is already a 1 in the second row. If Q is to be row complete, it cannot be 2, so it must be 0. But that means $Q_{2,3} = 2$, which is impossible because $Q_{1,3} = 2$. So there is no RCLS(3).

When trying to answer Judy's question for five quilters, the first thing to notice is that there are essentially only two ways that the first trade could go. Either the group could pass the quilts in one big cycle, or two people could trade and the other three could pass in a cycle. In each case, the initial passes determine the first two columns of the square. Since we have fixed these two columns, we can no longer rename the entries so that the first row assumes a particular sequence as we did in the case of 3.

- **Case 1:** Say the first trade involves everyone passing the quilts in one cycle. Without loss of generality, assume the first passing cycle is

$$0 \to 1 \to 2 \to 3 \to 4 \to 0.$$

The first two columns of the square are then determined, as shown below:

$A =$

0	1	?	?	?
1	2			
2	3			
3	4			
4	0			

Following the 1 in the first row, we have $A_{1,3} \in \{2, 3, 4\}$. It can't be that $A_{1,3} = 2$ because the sequence $(1, 2)$ appears in the second row. Say $A_{1,3} = 3$. Then $A_{1,4} \neq 4$ since $(3, 4)$ appears in the fourth row. So the first row must be $(0, 1, 3, 2, 4)$. Now consider the second row. It must be that $A_{2,3} \in \{0, 3, 4\}$. Since the sequence $(2, 3)$ appears in the third row and $(2, 4)$ appears in the first row, $A_{2,3} = 0$. Then $A_{2,4} = 4$, since 4 appears already in the fifth column, which means the second row is $(1, 2, 0, 4, 3)$; the square now looks like the following:

$A =$

0	1	3	2	4
1	2	0	4	3
2	3			
3	4			
4	0			

We now see that it is impossible to fill the third row subject to the constraints because $A_{3,3}$ must be 4 (since a 4 cannot appear in either of the last two columns), but then the sequence $(3, 4)$ appears twice in the square. Thus $A_{1,3} \neq 3$. By similar reasoning, we find that $A_{1,3} \neq 4$. Therefore the first pass cannot be in a single cycle.

- **Case 2:** Assume that on the first quilt pass, two people swap and the other three pass in a cycle. Without loss of generality, the first round of trades gives the following first two columns.

$A =$

0	1	?	?	?
1	0			
2	3			
3	4			
4	2			

Now we can rather quickly see that this will not work out. Notice that in each of the last three rows, we will need to place a 0 and a 1 in the last three columns. We can never put them next to each other because $(0, 1)$ appears in the first row and $(1, 0)$ appears in the second row. So the 0 and 1 will always have to fall in the third and fifth columns. But we have three rows where we need to place 0s, and two columns to place them, so that would force us to place two 0s in some column, which is not allowed. So this whole case falls apart.

Judy was correct that her circle could not be squared: since no RCLS(n) exists when $n = 3$ or 5, these sizes of quilting circles will never work. By Williams' construction, an RCLS(n) always exists when n is even, so these exchanges can always be arranged for even-sized quilting circles. Can we completely classify the possible sizes of quilting circles for which these exchanges are possible—that is, can we determine the values of n for which an RCLS(n) exists? Further, Williams' construction gives a rotational RCLS(n) for all even n. For which n do there exist rotational RCLS(n)? Equivalently, when will there be a single passing pattern that works for everyone?

Passing quilts on graphs

A natural, visual way to model the quilt exchange is by drawing the paths of the quilts in a complete digraph on n vertices. Each quilt follows a path in the digraph. These digraph paths are particularly useful in trying to find rotational RCLS. Each vertex is a member of the circle and each directed edge, called an arc, represents the action of passing the quilt from one member to another. Figure 3 shows a digraph representation of the RCLS(6) from Table 1.

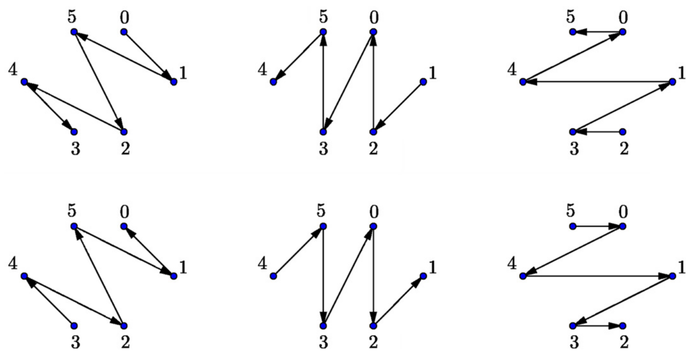


Figure 3 Digraph visual of the RCLS(6) in Table 1.

A Hamiltonian path in a digraph is an ordering of the vertices so that each vertex appears exactly once and every ordered pair of consecutive vertices in the sequence has a corresponding arc in the digraph. A Hamiltonian path would thus denote the path that a particular quilt takes through the quilting circle.

A *complete digraph* on n vertices, denoted D_n, connects each pair of vertices with two arcs, one in each direction. Creating n quilts in accordance with the quilting circle restrictions is equivalent to finding n arc-disjoint Hamiltonian paths in D_n such that no vertex appears in the same position of two different paths. We will call this last condition the *column latin condition*.

Since there are $n(n-1)$ arcs in D_n and each quilt would require $n-1$ arcs to be complete, every arc must be used. A collection of Hamiltonian paths that partition the arcs of D_n is called a *Hamiltonian path decomposition* of the digraph. We seek a Hamiltonian path decomposition of D_n with the column latin condition that each vertex appears exactly once in each path position. An example with $n = 4$ is shown in Figure 4.

As Figure 4 shows, having a decomposition into paths where each successive graph can be obtained by a cyclic rotation of the previous one is a convenient way to guarantee that the column latin condition is satisfied. We construct another example of this property on eight vertices, as demonstrated in Figure 5.

These paths are arc-disjoint because each arc in the first path was chosen to be a different length, where "length" means the number of clockwise steps around the circle we take before landing on the next adjacent vertex. Therefore, when we rotate the path no arc will be repeated. If we label the vertices of the graphs in Figure 5 with

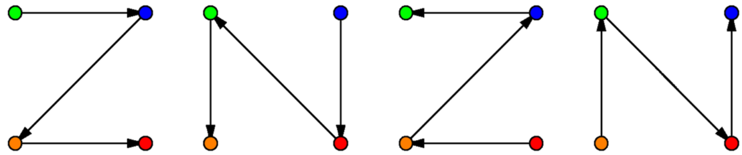


Figure 4 Decomposing D_4 into four arc-disjoint Hamiltonian paths with the column latin condition.

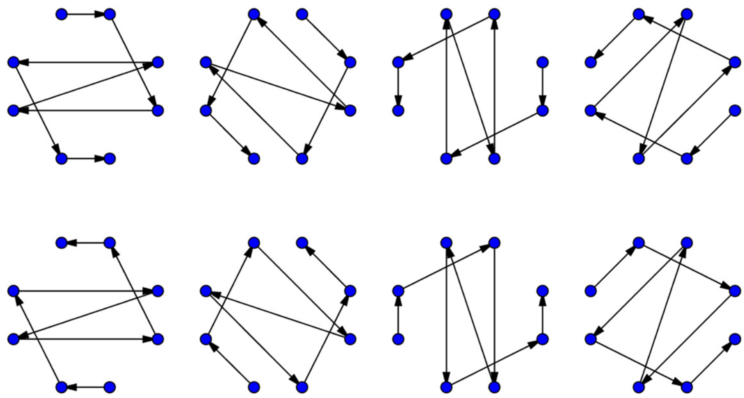


Figure 5 Decomposing D_8 into eight arc-disjoint Hamiltonian paths with the column latin condition.

the numbers 0, 1, . . . , 7 in the natural way, then each Hamiltonian path gives a row in an RCLS. The resulting square is shown in Table 2.

0	1	3	6	2	7	5	4
1	2	4	7	3	0	6	5
2	3	5	0	4	1	7	6
3	4	6	1	5	2	0	7
4	5	7	2	6	3	1	0
5	6	0	3	7	4	2	1
6	7	1	4	0	5	3	2
7	0	2	5	1	6	4	3

TABLE 2: An RCLS(8) resulting from the Hamiltonian path decomposition in Figure 5.

The cyclic rotation property of this Hamiltonian path decomposition results in identical successive differences (modulo n) in each row of the latin square, thus producing a rotational latin square. Finding a rotational RCLS(n) is equivalent to finding a single Hamiltonian path in D_n which is disjoint with each of its $n-1$ non-trivial rotations.

Rotational squares and triangular numbers

In this section we find the values of n for which a particular sequence of successive differences can form a rotational RCLS(n) with entries in $\{0, 1, \ldots, n-1\}$. Notice that not every sequence of successive differences will give a latin square at all. The sequence of successive differences $(d_1, d_2, ..., d_{n-1})$ will give a rotational latin square if and only if none of the partial sums $\sum_{i=1}^{k} d_i$ are congruent modulo n for $k = 1, \ldots, n-1$. To see this, suppose that the entry in the upper left of the square is 0. Then the first row of the latin square will be equivalent to

$$\left(0, d_1, d_1 + d_2, d_1 + d_2 + d_3, \ldots, \sum_{i=1}^{n-1} d_i\right) \mod n.$$

If two of the partial sums are equivalent modulo n, then the square will have repeated entries in the first row, so it cannot be a latin square. Conversely, suppose that the partial sums are all distinct modulo n, and assume that the first column of the square is $(0, 1, \ldots, n-1)$. Then each row has distinct entries modulo n, since the partial sums are distinct. Moreover, the columns also have distinct entries modulo n, since each column is an arithmetic progression with common difference one. Therefore the resulting square is a rotational latin square.

Further, note that a rotational latin square will be an RCLS exactly when the successive differences are all distinct. In terms of the digraph, the successive differences are all distinct if and only if the clockwise length of each arc in a path is different. If and only if this is the case, each path will be disjoint from its $n-1$ rotations in D_n. Consider the successive differences $(1, 2, 3, 4, ..., n-1)$, which play a special role in the examples in Figures 4 and 5. The partial sums for these successive differences are the first $n-1$ triangular numbers. The triangular number T_k is the sum of the consecutive integers from 0 to k.

Let $S = \{0, 1, \ldots, n-1\}$ be interpreted as the integers modulo n and be denoted $\mathbb{Z}/n\mathbb{Z}$. Let k be a natural number. We will show that, for certain $n = 2^k$, the successive differences $(1, 2, \ldots, n-1)$ give rise to a rotational RCLS(n) with entries from S. From the discussion above, we know that this construction results in a square with first row $(T_0, T_1, T_2, \ldots, T_{n-1}) \mod n$, and that the square will be a rotational RCLS(n) exactly when these triangular numbers are all distinct modulo n. This is the key to the proof of the following theorem.

Theorem 1. *The sequence of differences* $(1, 2, \ldots, (n-1))$ *results in a rotational RCLS if and only if n is a power of 2.*

To prove this result, we make use of the fact that $T_n = \sum_{i=0}^{n} i = \frac{(n+1)(n)}{2}$. We begin with a lemma.

Lemma 1. *Let T_i denote the i-th triangular number. If $n = 2^k$, then no two triangular numbers T_i and T_j with* $0 < i < j < n$ *are equivalent modulo n.*

This is given as an exercise in Donald Knuth's book *The Art of Computer Programming* [**14**], which we present here because it illustrates the remarkable power of data to reveal a path to proof. Two patterns become obvious from the data in Table 3.

First, we find pairs congruent mod 2^{k-1} within the first 2^k triangular numbers.

Claim 1. *If $i + j = 2^k - 1$, then $T_i \equiv T_j \mod 2^{k-1}$.*

Proof. Say $i + j = 2^k - 1$. Then we see that

$$T_j - T_i = (i+1) + (i+2) + \cdots + j = \frac{(i+1+j)(j-i)}{2}$$
$$= \frac{2^k(j-i)}{2} = 2^{k-1}(j-i).$$

Therefore $T_j - T_i \equiv 0 \mod 2^{k-1}$. ■

Now, we see an incongruence mod 2^k for these same numbers.

Claim 2. *If $i + j = 2^k - 1$, then $T_j \equiv T_i + 2^{k-1} \mod 2^k$.*

Proof. Assume that $i + j = 2^k - 1$. Then, as above, $T_j - T_i = 2^{k-1}(j - i)$. Since $i + j$ is odd, then $j - i$ is also odd, so $j - i = 2a + 1$ for some $a \in \mathbb{Z}$. Therefore

$$T_j - T_i = 2^{k-1}(2a + 1) = 2^k a + 2^{k-1}.$$

So $T_j - T_i \equiv 2^{k-1} \mod 2^k$. ■

Proof of Lemma 1. We proceed by induction. The base case follows from the data above. Assume that the theorem holds for $k = m - 1$. That means all T_i with $0 \leq i < 2^{m-1}$ are distinct modulo 2^{m-1}. By Claim 1, all T_j with $2^{m-1} \leq j < 2^m$ are also distinct modulo 2^{m-1}. Therefore each equivalence class modulo 2^{m-1} occurs for exactly one T_i with $0 \leq i < 2^{m-1}$ and exactly one T_j with $2^{m-1} \leq j < 2^m$. Therefore for all i, j with $0 \leq i < j < 2^m$, if $i + j \neq 2^m - 1$, then $T_i \not\equiv T_j \mod 2^{m-1}$, meaning $T_i \not\equiv T_j \mod 2^m$. However, if $i + j = 2^m - 1$, then Claim 2 implies that T_i and T_j must be distinct modulo 2^m. ■

A number that is the sum of at least two consecutive positive integers is called a *trapezoidal number* [7], or a *polite number* [1]. The former name reflects the fact that a trapezoidal number is a difference of two triangular numbers, which can be visualized as a trapezoidal array. In [7], it is proven that a number is trapezoidal if and only if it is not a power of 2. This result also appears in [10], and seems to have been rediscovered multiple times since then.

T_n	mod 2	mod 4	mod 8	mod 16
$T_0 = 0$	0	0	0	0
$T_1 = 1$	1	1	1	1
$T_2 = 3$	1	3	3	3
$T_3 = 6$	0	2	6	6
$T_4 = 10$		2	2	10
$T_5 = 15$		3	7	15
$T_6 = 21$		1	5	5
$T_7 = 28$		0	4	12
$T_8 = 36$			4	4
$T_9 = 45$			5	13
$T_{10} = 55$			7	7
$T_{11} = 66$			2	2
$T_{12} = 78$			6	14
$T_{13} = 91$			3	11
$T_{14} = 105$			1	9
$T_{15} = 120$			0	8

TABLE 3: Data to see that, if $n = 2^k$ and $0 < i < j < n$, then $T_i \not\equiv T_j \mod n$.

0	1	3	2
1	2	0	3
2	3	1	0
3	0	2	1

Figure 6 A rotational RCLS(4) with successive differences $(1, 2, 3)$.

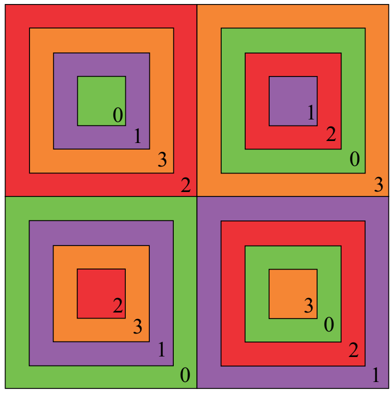


Figure 7 Four quilts associated to the RCLS(4) in Figure 6.

Theorem 2 (Gamer et al. [7]). *All positive integers except the powers of* 2 *are trapezoidal.*

We show that if a number is trapezoidal, then $(1, 2, \ldots, n-1)$, the sequence of differences, cannot result in an RCLS, because there is a repeat among the congruence classes of the triangular numbers modulo n.

Lemma 2. *If the number n is trapezoidal, then there exists i, j such that $0 < i < j < n$, and $T_i \equiv T_j \mod n$.*

Proof. Assume n is trapezoidal, so there are positive integers i and l such that

$$n = (i+1) + (i+2) + \cdots + (i+l),$$

and set $j = i + l$. Notice that the sum can be written as the difference of two triangular numbers:

$$(i+1) + (i+2) + \cdots + (i+l) = T_{i+l} - T_i = T_j - T_i$$

Therefore $n = T_j - T_i$, where $0 < i < j < n$, and the equivalence $T_j \equiv T_i \mod n$ holds. ■

We then have Theorem 1 as a corollary.

Theorem 3. *The sequence of differences $(1, 2, \ldots, (n-1))$ results in a rotational RCLS if and only if n is a power of 2.*

Proof. Recall that the sequence of differences $(1, 2, \ldots, (n-1))$, with first entry 0 lead to the row $(T_0, T_1, \ldots, T_{n-1}) \mod n$. By Lemma 1, we have that none of these entries are equivalent modulo n when n is a power of 2, so all entries in this row are distinct. By Theorem 2 and Lemma 2, we see that if n is not a power of 2, then there will be repeated entries in this row. Therefore the rotational construction described at the beginning of the section results in a rotational latin square if and only if n is a power of 2. Because the sequence of differences are all distinct, this rotational latin square is row complete. ■

Square groups

The search for rotational latin squares leads us to another area of mathematics: group theory. Let G be a finite group and $(g_1, g_2, \ldots, g_n)$ be some ordering of its elements. The Cayley, or group operation, table of G is the $n \times n$ array C with $C_{i,j} = g_i g_j$. The existence of inverses in groups implies that for all $a, b, c \in G$, if $ab = cb$, then $a = c$, and if $ab = ac$, then $b = c$. Therefore the Cayley table of a group must form a latin square. Is every latin square the (relabeled) Cayley table of some group? Not necessarily, as illustrated by

0	1	2
2	0	1
1	2	0

,

since $\mathbb{Z}/3\mathbb{Z}$ is the only group of order 3 and $\mathbb{Z}/3\mathbb{Z}$ is abelian, thus its Cayley table is symmetrical across the diagonal. However, in this case, the rows can easily be rearranged to yield the Cayley table of $\mathbb{Z}/3\mathbb{Z}$. There are latin squares whose rows cannot be rearranged to yield the Cayley table of a group, such as

0	1	2	3	4
1	2	4	0	3
2	3	1	4	0
3	4	0	1	2
4	0	3	2	1

.

Again, there is only one group of order 5, namely $\mathbb{Z}/5\mathbb{Z}$ with the operation addition, which is an abelian group. In this case there is no reordering of the rows that gives symmetry across the diagonal. In short, group tables always yield latin squares, but not all latin squares arise from group tables.

Say that a row complete latin square arises from a group. What restrictions would row completeness place on the group structure? It would have to be possible to order the group elements $\{g_1, g_2, \ldots, g_n\}$ so that $(g_i g_k, g_i g_{k+1}) = (g_j g_m, g_j g_{m+1})$ implies $i = j$ and $k = m$ for all $1 \leq i, j \leq n$ and all $1 \leq k, m \leq n - 1$. So row completeness is a statement about how group elements can be ordered. This leads us to consider sequenceable groups.

A finite group G of order $n > 1$ is said to be *sequenceable* if its elements can be listed in some order $(g_1, g_2, \ldots, g_n)$ such that all products $a_i = g_1 g_2 \ldots g_i$ are distinct. Gordon proved that if G is sequenceable, then G gives rise to an RCLS as follows [**9**]. Define $c_{i,j} = a_i^{-1} a_j$. Then the rows of the array $[c_{i,j}]$ can be reordered to form the Cayley table for G. See Table 4.

	a_1	a_2	$\cdots$	a_{n-1}	a_n
a_1^{-1}	e	$a_1^{-1}a_2$	$\cdots$	$a_1^{-1}a_{n-1}$	$a_1^{-1}a_n$
a_2^{-1}	$a_2^{-1}a_1$	e	$\cdots$	$a_2^{-1}a_{n-1}$	$a_2^{-1}a_n$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_{n-1}^{-1}	$a_{n-1}^{-1}a_1$	$a_{n-1}^{-1}a_2$	$\cdots$	e	$a_{n-1}^{-1}a_n$
a_n^{-1}	$a_n^{-1}a_1$	$a_n^{-1}a_2$	$\cdots$	$a_n^{-1}a_{n-1}$	e

TABLE 4: The RCLS formed from a sequenceable group.

What groups are sequenceable? From Theorem 1, we can see that $\mathbb{Z}/2^k\mathbb{Z}$ is sequenceable, with the sequencing $(0, 1, \ldots, 2^k - 1)$.

Theorem 4 (Gordon [**9**]). *A finite abelian group G is sequenceable if and only if $G = \mathbb{Z}/2^k\mathbb{Z} \times B$, where $k \geq 1$ and B is of odd order.*

For non-abelian groups, the question is not resolved. Several things are known, including that the dihedral groups D_3 (order 6) and D_4 (order 8), and the group of quaternions Q (order 8) are not sequenceable. It has been proven that D_r is sequenceable for all $r > 4$. Are there sequenceable groups of odd order? Yes—a few infinite classes have been proven to be sequenceable [**13**, **19**]. Keedwell [**13**] conjectured that in fact all non-abelian groups of order greater than 8 are sequenceable, a conjecture checked by Anderson for orders 9 through 32 [**2**, **3**].

Gordon's clever construction results in a latin square that is also column complete. At the time, it seemed possible that all row complete latin squares were derived from sequenceable groups, and that perhaps the rows of any RCLS could be reordered to create a column complete RCLS. Denes and Keedwell asked whether this was the case in their text [**6**].

However, in 1974, Owens proved not every RCLS comes from a sequenceable group, and that not every RCLS could be rearranged to form a column complete latin square [**18**]. So while the existence of a sequenceable group of order n implies the existence of RCLS(n) for all even values and some odd n, results about sequencable groups do not tell us any n for which it is **not** possible to construct an RCLS of order n.

Returning to rotational RCLS, we can now see that creating a rotational RCLS(n) with the sequence of successive differences $(d_1, d_2, \ldots, d_{n-1})$ is equivalent to the group $\mathbb{Z}/n\mathbb{Z}$ admitting the sequencing $(0, d_1, d_2, \ldots, d_{n-1})$. Thus a rotational RCLS(n) exists if and only if the group $\mathbb{Z}/n\mathbb{Z}$ is sequenceable, thus (by Theorem 4) if and only if n is even.

A Composite Breakthrough

In 1968, Mendelsohn discovered an RCLS of size 21, the first known RCLS of odd order $n > 1$ [**17**]. Subsequently, RCLS of orders 9, 15, 25, 27, 33, 39, and 55 were discovered. Many were created using the concept of generating arrays, introduced in 1980 by Archdeacon et al. [**4**] and adapted by others including Higham [**11**]. Let q be an odd prime power and m be any natural number, and let $n = mq$. Though a full description can be found in [**4**], the basic idea of a generating array is to build an RCLS(n) from a $q \times n$ array A with entries in $\mathbb{F}_q \times \mathbb{Z}/m\mathbb{Z}$ satisfying certain conditions.

Letting $q = m = 3$, we have the following example of a generating array A and the resulting latin square L [4]:

$A =$

(0, 0)	(1, 0)	(2, 0)	(0, 1)	(1, 2)	(2, 1)	(1, 1)	(2, 2)	(0, 2)
(1, 0)	(0, 1)	(0, 0)	(2, 1)	(2, 0)	(0, 2)	(2, 2)	(1, 1)	(1, 2)
(2, 0)	(2, 1)	(1, 2)	(1, 1)	(0, 1)	(1, 0)	(0, 2)	(0, 0)	(2, 2)

,

$L =$

0	1	2	3	7	5	4	8	6
3	4	5	6	1	8	7	2	0
6	7	8	0	4	2	1	5	3
1	3	0	5	2	6	8	4	7
4	6	3	8	5	0	2	5	1
7	0	6	2	8	3	5	1	4
2	5	7	4	3	1	6	0	8
5	8	1	7	6	4	0	3	2
8	2	3	1	0	7	3	6	5

.

The last major theoretical development in the field dates from 1996, when Higham developed methods for creating generating arrays for all odd composite $n > 9$, thus proving that RCLS(n) exist for all composite n [11]. It is known that RCLS of orders 3, 5, and 7 do not exist, but existence is unknown for any larger primes.

The unknown

We still would like to know exactly what sizes of quilting circles allow this type of exchange. The preceding section shows that all composite size circles will work. The most prominent open question in the area is whether any primes will work.

Question 1. *Does there exist an RCLS(11)? More generally, are there any RCLS of prime order?*

Given that there are, up to equivalence, approximately 2×10^{24} distinct classes of latin squares of order 11 [12], and that this equivalence does not preserve row completeness, this is a daunting computational problem. In fact, a more reasonable estimate of the difficulty of the problem may be given by the total number of reduced latin squares of order 11, approximately 5.4×10^{34} [16]. Significant computational power or new techniques may be required to answer this question for $n = 11$.

In the section on square groups, we saw that a rotational RCLS(n) exists exactly when n is even. When $n = 2^k$, we have seen that $(1, 2, \ldots, n-1)$ and $(1, -2, 3, -4, \ldots, n-1)$ are both valid sequences of successive differences, corresponding to different sequencings of the group $\mathbb{Z}/n\mathbb{Z}$. This illustrates that it is possible to have more than one sequence of differences that yield a rotational RCLS for a given n.

Question 2. *How many different sequences of differences yield rotational RCLS(n) for a given n?*

Rotational RCLS would give rise to exceptionally simple passing patterns for quilts, where each person follows the same pattern in the exchange. This question is certainly more accessible, and the study of sequencings of cyclic groups may lead to new rotational RCLS. Answering this question is equivalent to determining all sequencings for cyclic groups, an open question. Gilbert [8] gives several infinite classes of sequencings, and determines all sequencings for $n \leq 10$, but there may be unknown sequencings for larger cyclic groups.

As a final direction for further inquiry, we introduce quiltdoku. Consider a scenario where we would like particular quilts to be in certain places at given times. Is it possible to devise a passing pattern that will satisfy these constraints? This problem can be phrased as a puzzle as follows: given a partially filled in grid, is it possible to fill the remainder of the boxes to create an RCLS? Consider the following examples:

0			
	3		
		1	
			1

0	1	2	3	4	5
	2				
		4			
			4		
				2	
5					0

To satisfy the puzzlers of the world, we might like to have a unique solution to a particular partially completed grid. In creating these challenging but feasible quiltdoku puzzles, we like to know the smallest number of clues we can give while still arriving at a unique solution. The answer to the corresponding question for Sudoku (a 9×9 latin square with the additional constraint that every digit appears in each of the 3×3 boxes obtained by dividing the grid into thirds horizontally and vertically) was found to be 17 [15], but since quiltdoku puzzles could exist for n of varying sizes, the answer will not be a single number nor perhaps even a single formula. An immediate observation is that for $n \times n$ puzzles, if there are two distinct rows that do not contain any clues, the puzzle is not uniquely completable, because in the solution the two rows can be interchanged. This provides us with a lower bound on the number of clues needed, $n - 1$.

Question 3. *What partial grids can be uniquely completed to RCLS? What is the smallest number of clues necessary for a unique solution to a quiltdoku?*

Acknowledgment This work partially supported by NSA grant #H98230-16-1-0300. We would like to thank Judy Gilmore for the question that started this quest, as well as John Watkins and anonymous referees for their careful reading and excellent suggestions.

REFERENCES

[1] Adams, K. (1993). How polite is x? *Math. Gazette*. 77 (478): 79–80.

[2] Anderson, B.A. (1987). A fast method for sequencing low order non-abelian groups. *Ann. Discrete Math.* 34: 27–41.

[3] Anderson, B. A. (1987). S_5, A_5 and all nonabelian groups of order 32 are sequenceable. *Congr. Numer.* 58: 53–68.

[4] Archdeacon, D. S., Dinitz, J. H., Stinson, D. R., Tillson, T. W. (1980). Some new row-complete latin squares. *J. Combin. Theory Ser. A*. 29 (3): 395–398.

[5] Ball, R. D. (1997). Incomplete block designs for the minimisation of order and carry-over effects in sensory analysis. *Food Qual. Preference*. 8 (2): 111–118.

[6] Dénes, J., Keedwell, A. D. (1991). *Latin Squares: New Developments in the Theory and Applications*. Amsterdam: New Holland.

[7] Gamer, C., Roeder, D. W., Watkins, J. J. (1985). Trapezoidal numbers. *Math. Mag.* 58 (2): 108–110.

[8] Gilbert, E. N. (1965). Latin squares which contain no repeated digrams. *SIAM Rev.* 7 (2): 189–198.

[9] Gordon, B. (1961). Sequences in groups with distinct partial products. *Pacific J. Math.* 11: 1309–1313.

[10] Griggs, T. S. (1991). Impolite numbers. *Math. Gazette*. 75 (474): 442–443.

[11] Higham, J. (1998). Row-complete latin squares of every composite order exist. *J. Combin. Designs*. 6 (1): 63–77.

[12] Hulpke, A., Kaski, P., Ostergard, P. (2011). The number of latin squares of order 11. *Math. Comput*. 80 (274): 1197–1219.

[13] Keedwell, A. D. (1981). On the sequenceability of non-abelian groups of order pq. *Discret. Math.* 37 (2/3): 203–216.

[14] Knuth, D. E. (1998). *The Art of Computer Programming, Volume 3: Sorting and Searching*. Redwood City, CA: Addison Wesley Longman.

[15] McGuire, G., Tugemann, B., Civario, G. (2014). There Is no 16-clue Sudoku: Solving the Sudoku minimum number of clues problem via hitting set enumeration. *Exp. Math.* 23 (2): 190–217.

[16] McKay, B. D., Wanless, I. M. (2005). On the number of latin squares. *Ann. Combin*. 9 (3): 335–344.

[17] Mendelsohn, N. S. (1968). Hamiltonian decomposition of the complete directed n-graph. In: P. Erdös, G. Katona, eds. *Theory of Graphs (Proceedings Colloq. Tihanny, 1966)*. Academic Press: New York, pp. 237–241.

[18] Owens, P. J. (1976). Solutions to two problems of Dénes and Keedwell on row complete latin squares. *J. Combin. Theory, Ser. A*. 21 (3): 299–308.

[19] Wang, L. L. (1973). A test for the sequencing of a class of finite groups with two generators. *Amer. Math. Soc. Notices.* 20: 73T–A275.

[20] Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Austral. J. Scient. Res., Ser. A: Phys. Sci.* 2: 149–168.

Summary. In a certain type of round robin quilting circle, each person starts their own quilt by making a square of some color or design. Everybody in the circle then passes their quilt to someone else, who adds on to the outside, and then passes it along again, until each person has contributed to every quilt. Passing the quilts in a literal circle would work, but is not ideal from a social or design perspective, since person A's work is always followed by that of person B, and A always interacts with B during the passing. Can we create a scheme so that each person passes a quilt to every other person in the circle? This situation can be modeled mathematically using combinatorial objects called row complete Latin squares. This paper explores the known and unknown aspects of this colorful problem.

KATIE HAYMAKER (MR Author ID: 1048523) obtained a Ph.D. in mathematics from the University of Nebraska–Lincoln in 2014. She is an assistant professor in the Department of Mathematics and Statistics at Villanova University. Dr. Haymaker's research focuses on coding theory and applied discrete mathematics. She is grateful that Dr. Malmskog included her in discussions of this fascinating problem, which has led to many exciting discoveries and re-discoveries.

BETH MALMSKOG (MR Author ID: 897266) wishes more of her friends would turn their hobbies into math problems for her to work on. Dr. Malmskog received her PhD from Colorado State University in 2011. Most of her research is in number theory. She is now an assistant professor in the Department of Mathematics and Computer Science at Colorado College.

The Settlers of "Catanbinatorics"

JATHAN AUSTIN
Salisbury University
Salisbury, MD 21801
jwaustin@salisbury.edu

BRIAN G. KRONENTHAL
Kutztown University of Pennsylvania
Kutztown, PA 19530
kronenthal@kutztown.edu

SUSANNA MOLITORIS MILLER
Kennesaw State University
Kennesaw, GA 30144
smolitor@kennesaw.edu

Catan (formerly known as *The Settlers of Catan*) is a board game based on property development and resource trading. Like many other games, *Catan* contains opportunities for the application of game theory, probability, and statistics (see, e.g., [1]). However, some games also provide interesting contexts for exploring combinatorics (see, e.g., [3]). *Catan* is one such game due to its game board design which allows players to "construct" a new board every time they play by randomly arranging nineteen hexagonal tiles, eighteen number tokens, and nine port (harbor) markers according to a set of given parameters. To many, this leads to seemingly endless possible boards, but a mathematician will likely raise the "Catanbinatorics" question of exactly how many possible boards exist. In this paper we use basic combinatorial techniques to explore this question. We also address two related counting problems by focusing on parts of the game board design. The first reconsiders the way in which we count the arrangements of number tokens based on their role in the game. The second explores two methods of counting non-equivalent ways to arrange only the resource tiles. One might expect that no longer considering the number tokens and ports would simplify calculations, however, removing these components surprisingly makes the problem more complex (and interesting!) to solve.

A Brief History of *Catan*

Catan is an award-winning, internationally popular, easy-to-learn strategy board game which has been credited with revolutionizing the board game industry [7]. Since its introduction in Germany in 1995, Klaus Teuber's innovative game has received numerous awards including, but not limited to, Spiel des Jahres Game of the Year (1995), Meeples Choice Award (1995), Games Magazine Hall of Fame (2005), and GamesCon Vegas Game of the Century (2015) [2]. As of 2015, *Catan* has sold over 22 million copies and has been translated into over 30 different languages [9]. It has inspired several expansions and themed game variations, as well as several digital adaptations for platforms such as Microsoft, Nintendo, Xbox, iPod/iPhone, and Facebook [2]. The seasoned *Catan* player may even notice some subtle differences between editions of the base game. We will not explore these expansions and variations.

The complexities of *Catan* as a strategy game have received attention in both professional and recreational domains. Computer scientists have praised *Catan* as a scenario ripe with potential for artificial intelligence and programming analysis (see, e.g.,

Math. Mag. **92** (2019) 187–198. doi:10.1080/0025570X.2019.1561096
MSC: Primary 05A15, Secondary 20B30
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

[6, 11, 13]). Mathematicians have highlighted how one might mathematize the choices made during initial settlement placement by using statistics and expected value to assign values to potential settlement locations based on players' individual strategies [1]. *Catan* has also received significant attention in various amateur circles via blog posts and other unreviewed works. Several of these focus on counting problems related to *Catan*, including how many distinct possible boards exist (see, e.g., [8, 10, 12]). Many present correct information, and furthermore some discuss approaches similar to what appears in this paper. However, the mathematics presented here was developed independently. Due to the general unreliability of unreviewed information, we assert the value of an authoritative and mathematically accurate exploration that is widely accessible while still substantive and interesting. We hope the reader will find that this paper satisfies these goals.

Board Construction

According to the *Catan* game rules, the board is assembled in three stages: the resource tiles, the number tokens, and the ports. The first step is positioning the nineteen hexagonal resource tiles in a larger roughly hexagonal configuration shown in Figure 1. These tiles designate which resources will be produced by each location on the board. There are four lumber tiles, four grain tiles, four wool tiles, three brick tiles, three ore tiles, and one desert tile which does not produce any resources.

Next the number tokens are arranged, one per hexagon resource tile with the exception of the desert. The number tokens are labeled "2" through "12" (excluding "7"), with one "2," one "12," and two of each for numbers "3" through "11" (except "7"). These tokens are placed in one-to-one correspondence with the resource tiles and dictate when each resource will be produced during the game; at the start of each turn, a player rolls a pair of dice and the resource tiles whose label matches the roll will produce resources for any player who has a settlement adjacent to the tile. The game rules dictate that tokens with red numbers (labeled "6" or "8") cannot be next to one another; however, for the purposes of this article, we are opting to ignore this restriction in favor of an entirely random set-up design.

The final phase of board construction involves placing ports at different designated locations around the larger hexagonal configuration, also shown in Figure 1. The ports allow players to trade two of a specified resource type for one of any other or to trade three of any common type for one of any other. There is one port for each of the five resources (lumber, brick, wool, grain, and ore), and four ports which allow any resource to be traded at the reduced three-for-one rate.

How Many Boards?

One's first instinct when counting the number of boards may be to consider it as no more than a relatively straight-forward combinatorial problem for permutations with repeated elements, similar to counting the number of possible arrangements of the letters in the word MISSISSIPPI. We begin with the resource tiles, number tokens, and ports, and then account for equivalent boards under symmetries.

In laying out the resource tiles we begin with the 19 tile locations and choose four for the lumber, four for the grain, four for the wool, three for the ore, and three for the brick, with the remaining spot designated as the desert. So the number of ways to arrange just the resource tiles would be

$$\binom{19}{4,4,4,3,3,1} = \binom{19}{4}\binom{15}{4}\binom{11}{4}\binom{7}{3}\binom{4}{3} = 244{,}432{,}188{,}000. \qquad (1)$$

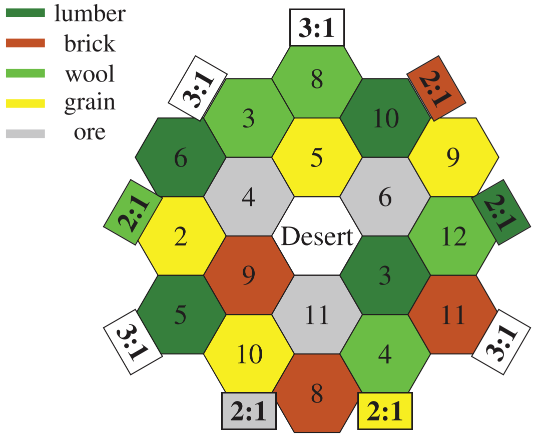


Figure 1 A sample board.

Then, adding in the number tokens would require selecting two of the eighteen non-desert spots for each number token from "3" to "11," excluding "7," and then choosing one of the remaining two spots for the "2" and the other by default for the "12." So the number of ways to place the number tokens would be

$$\binom{18}{2,2,2,2,2,2,2,2,1,1} = \binom{18}{2}\binom{16}{2}\binom{14}{2}\binom{12}{2}\binom{10}{2}\binom{8}{2}\binom{6}{2}\binom{4}{2}\binom{2}{1}$$
$$= 25{,}009{,}272{,}288{,}000.$$

Adding the ports is a much simpler process because we would simply choose four locations for the three-for-one ports and then distribute the five resource-specific ports among the remaining five positions, for a total of

$$\binom{9}{4} \cdot 5! = 15{,}120.$$

To obtain the total number of possible boards, we multiply the number of arrangements for each of these three components, i.e., the resource tiles, number tokens, and ports, to get more than 9.2429635×10^{28} possible configurations of the resources numbers and ports. While this may seem like the final number of boards, there is one more factor which must be taken into account. Because of the way *Catan* is played, the structure of the game board depends only on how various elements of the board are arranged relative to one another. The general arrangement of resource tiles, number tokens, and ports has $(120n)^\circ$ rotational symmetry for $n = 0, 1, 2$ and three lines of symmetry; see Figure 2. Any such rotation or reflection of a given board will create a new configuration while maintaining all salient adjacencies among board elements, and therefore can be thought of as equivalent to the original board. So for any given board there are five other equivalent boards as related by possible reflections, rotations, and combinations thereof. So we must divide our previous total number of configurations by six in order to account for the six equivalent versions of the same board. This leaves us with more than 1.5404939×10^{28} boards. But this is not the end; there are a few more im*port*ant "Catanbinatorial" questions to consider.

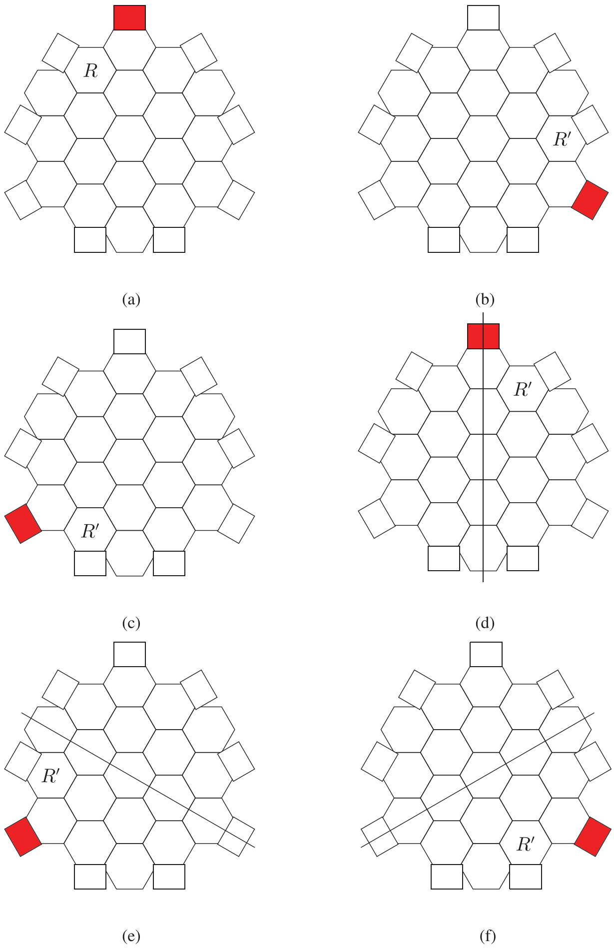


Figure 2 This figure contains images of six equivalent boards; they are simplified in that only one port and one resource tile are labeled to make the equivalences more visible. If (a) is considered the original board, (b) can be obtained by rotating (a) 120° clockwise, (c) can be obtained by rotating (a) 240° clockwise, and (d), (e), and (f) can each be obtained by reflecting (a) across the line shown on each respective board.

Equivalence among number token configurations

The strategy for counting the arrangements of the number tokens provided earlier can also be refined for equivalent configurations based on how the number tokens act during game play. The primary purpose of the numbers involves connecting the production of resources to the roll of a pair of standard six-sided dice. Each turn begins with a dice roll; resources are then produced by the resource tiles which have number tokens that match the number produced by the roll and are collected by any player(s) who have a settlement adjacent to the producing resource tiles. Because of this function, one may wish to think of the number tokens based on their probability of being rolled rather than the actual number printed on each. For example, a “6” and an “8” are essentially the same because they are equally likely to be rolled. Under this assumption there are two number tokens with probability $1/36$, and four each of tokens with probabilities $2/36$, $3/36$, $4/35$, and $5/36$. So instead of placing two “6” tokens and two “8” tokens, one can imagine distributing four tokens with probability 5/36.

This would reduce our calculation to

$$\binom{18}{4,4,4,4,2} = \binom{18}{4}\binom{14}{4}\binom{10}{4}\binom{6}{4} = 9{,}648{,}639{,}000.$$

This result is significantly smaller than our original estimate which contained nearly 25 trillion more possibilities. However, such groupings may seem a bit hasty to the seasoned *Catan* payer. Although in the long term a 6 and an 8 are equally likely to be rolled, the game has a much different feel depending on if your settlements are adjacent to duplicate number tokens such as two tokens labeled “6” or diversified number tokens such as a “6” and an “8.” In order to take this into account and still consider duplicate boards we consider only the possibility of switching *pairs* of numbers, such as switching the two “6” tokens with the two “8” tokens. Because there are five pairs of numbers with the same probability of being rolled, we can simply divide the original calculation by 2^5, one 2 for each pair that could be switched. This brings our total to

$$\frac{\binom{18}{2,2,2,2,2,2,2,2,1,1}}{2^5} = \frac{\binom{18}{2}\binom{16}{2}\binom{14}{2}\binom{12}{2}\binom{10}{2}\binom{8}{2}\binom{6}{2}\binom{4}{2}\binom{2}{1}}{2^5} = 781{,}539{,}759{,}000.$$

Replacing the original calculation for number tokens with this equivalent one would again reduce the total number of boards to more than 4.8140434×10^{26} possible boards. This is still a lot of boards. If we counted one possible board every second of every day for 365 days a year, it would still take over 1.5×10^{19} years to go through them all!

Counting Resource Configurations

In the following section, we explain two ways to count the total number of possible configurations of the 19 resource tiles alone, without considering the number tokens or ports. Why isn’t the answer 244,432,188,000 as calculated earlier in equation (1)? If we are only placing the resource tiles, we actually have a significantly more complicated system of symmetries to explore. When considering complete boards, the presence of the number tokens and ports eliminates these symmetries. Thus, this work must be considered as its own problem and cannot inform the board counting argument. The reader may wish to pause while admiring this mathematical oddity: one might expect that removing the number tokens and ports from consideration would make calculations easier. However, this simplification surprisingly makes the problem more complex to solve.

Hence, our main objective in this section is to account for this more complicated system of symmetries of resource tile configurations. We do so in two ways. The first uses a simple and readily accessible direct approach without any heavy machinery. The second approach is more elegant and makes use of abstract algebra.

First, we note that no nontrivial rotation (less than 360°) of a configuration can ever produce itself. Indeed, there are no fixed configurations under 60° or 300° rotations because there are not six copies of any single resource (see Figure 3a). Similarly, there are no fixed configurations under 120° or 240° rotations because there are not six sets of three like resources (see Figure 3b). Finally, there are no fixed configurations under a 180° rotation because there are not nine pairs of like resources (see Figure 3c).

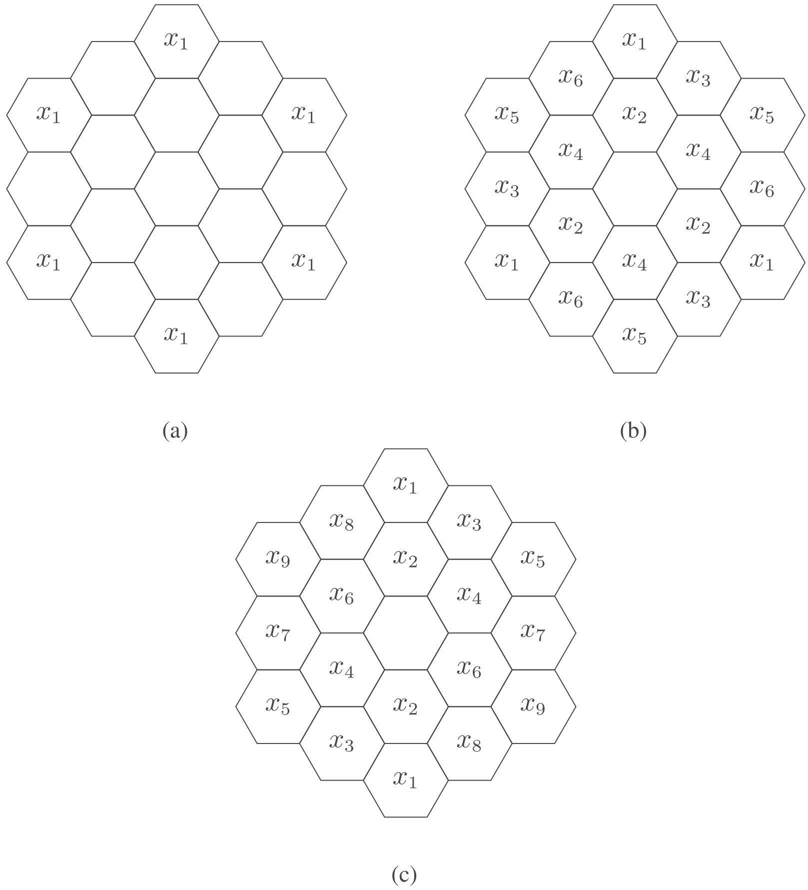


Figure 3 Nontrivial rotations of less than 360° do not fix any configuration.

A Direct Approach We begin by placing the desert; there are four choices, up to rotational and reflectional symmetry. For further explanation on type a and type b lines of symmetry, see Figures 4 and 5, respectively.

- **Case A:** The desert lies in the outer ring on a type a line of symmetry.
- **Case B:** The desert lies in the outer ring on a type b line of symmetry.
- **Case C:** The desert lies in the inner ring.
- **Case D:** The desert is the middle tile.

We begin with **case A**; without loss of generality, suppose the desert is placed in the uppermost location as shown in Figure 4.

Let T denote the set of all such configurations. Now,

$$|T| = \frac{18!}{4!4!4!3!3!} = 12{,}864{,}852{,}000.$$

Observe that any configuration that is NOT symmetric across line a will actually be counted twice: once for each of the two equivalent configurations. Denote by S_a the set of configurations that have reflectional symmetry across line a. Note that in all such configurations, the pair of resources placed in locations labeled $x_1, x_2, \ldots, x_7$

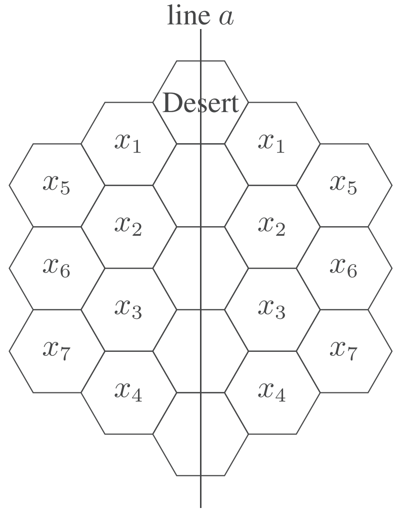


Figure 4 A case A configuration with symmetry across line a.

in Figure 4 must be the same. As there are eight available pairs (two pairs of wool tiles, two pairs of lumber, two pairs of grain, one pair of brick, and one pair of ore) in addition to a spare brick and ore, we proceed based on which of the eight pairs is not chosen.

If we leave out a pair of wool, lumber, or grain, then there are $\binom{3}{1} \cdot \frac{7!}{2!2!}$ ways to place the pairs because there are $\binom{3}{1}$ ways to choose which pair to exclude, and $\frac{7!}{2!2!}$ ways to place the 7 remaining pairs, dividing by 2!2! to account for the fact that there are two identical pairs which may each be interchanged without changing the configuration. We then must multiply by $\binom{4}{2} \cdot 2$, the number of ways to place the remaining tiles (the excluded pair of tiles, one brick, and one ore). Similarly, if we leave out a pair of brick or ore, then there are $\binom{2}{1} \cdot \frac{7!}{2!2!2!} \cdot \binom{4}{1}$ configurations. Hence, we have

$$|S_a| = \binom{3}{1} \cdot \frac{7!}{2!2!} \cdot \binom{4}{2} \cdot 2 + \binom{2}{1} \cdot \frac{7!}{2!2!2!} \cdot \binom{4}{1} = 50{,}400.$$

Therefore, since each configuration in $T \setminus S_a$ is double counted, the total number of distinct case A configurations up to symmetry is

$$\frac{|T| - |S_a|}{2} + |S_a| = \frac{|T| + |S_a|}{2} = 6{,}432{,}451{,}200.$$

We next consider **case B**; without loss of generality, suppose the desert is placed as shown in Figure 5.

As in the previous case, let T denote the set of all such configurations, double counting those that are not symmetric across line b; once again,

$$|T| = \frac{18!}{4!4!4!3!3!} = 12{,}864{,}852{,}000.$$

We let S_b denote the set of configurations that have reflectional symmetry across line b. This time, all eight pairs must be placed as illustrated in Figure 5, followed by the left-over brick and ore, and so

$$|S_b| = \frac{8!}{2!2!2!} \cdot 2 = 10{,}080.$$

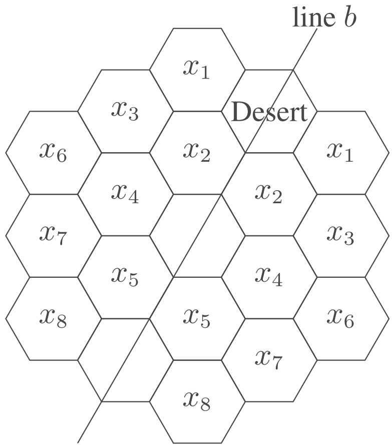


Figure 5 A case B configuration with symmetry across line b.

Hence, the total number of distinct case B configurations up to symmetry is

$$\frac{|T| - |S_b|}{2} + |S_b| = \frac{|T| + |S_b|}{2} = 6{,}432{,}431{,}040.$$

As can be seen by comparing Figure 6 to Figure 4, the number of **case C** configurations is equal to the number of case A configurations, namely 6,432,451,200.

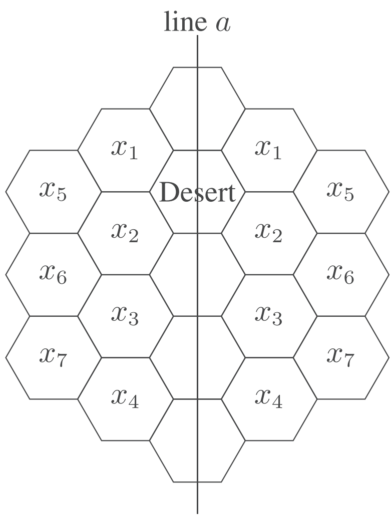


Figure 6 A case C configuration with symmetry across line a.

Finally, we consider **case D**. This case requires a bit more care due to the central location of the desert. It appears at first glance that we must consider rotational symmetry, but we already explained why this type of symmetry is impossible (see Figure 3 and the discussion immediately preceeding the "A Direct Approach" section). Therefore, we need only account for three potential lines of symmetry of type a and three potential lines of symmetry of type b. However, no configuration can be simultaneously fixed by a reflection of type a and a reflection of type b because this would either require six copies of a single resource (see Figure 7a) or three sets of four like resources and three additional pairs of like resources (see Figure 7b).

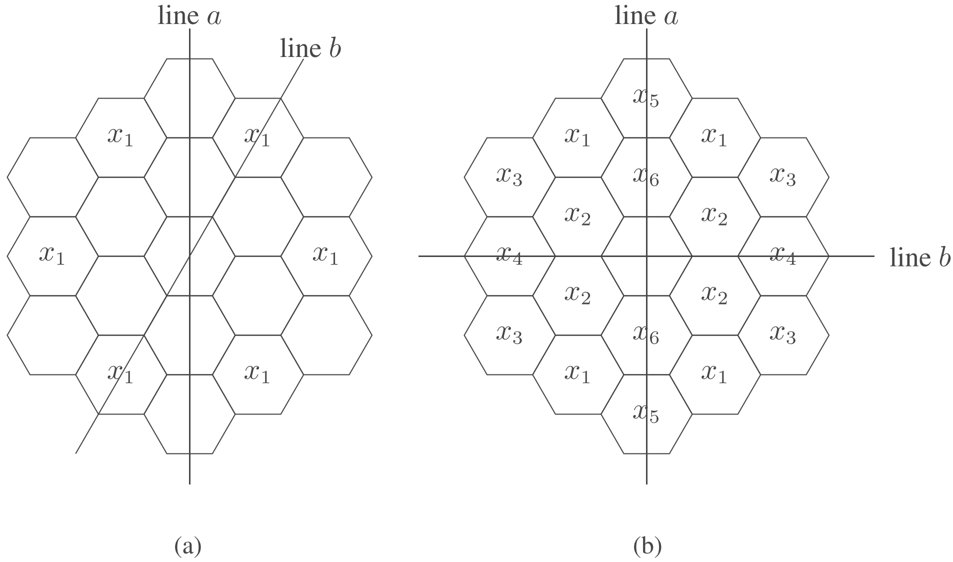


Figure 7 No configuration can be reflected onto itself using both type a and type b reflections.

Hence, the total number of distinct configurations up to symmetry is

$$\frac{1}{6}\cdot\left(\frac{|T|-3\left(|S_a|+|S_b|\right)}{2}+3\left(|S_a|+|S_b|\right)\right)=\frac{|T|+3\left(|S_a|+|S_b|\right)}{12}$$
$$=1{,}072{,}086{,}120,$$

where we divided by six because each configuration will be counted six times, one for each of the six possible rotations of the board.

Combining all cases, the total number of configurations up to symmetry is

$$6{,}432{,}451{,}200+6{,}432{,}431{,}040+6{,}432{,}451{,}200+1{,}072{,}086{,}120=20{,}369{,}419{,}560.$$

"Ore" would you prefer a more elegant approach?

A More Elegant Approach In this alternative approach, we will use Burnside's lemma to simplify our counting problem. We will need the following concepts.

Definition. Let G be a group of permutations on a set S (in other words, each element of G is a bijection $\phi : S \to S$). For any ϕ in G, define

$$\text{fix}(\phi)=\{i \in S | \phi(i)=i\}.$$

In other words, fix(ϕ) is the set of all elements of S that are fixed by ϕ.

Burnside's lemma is a statement about orbits. Again, let G be a group of permutations on a set S. Then for any $s \in S$, the orbit of G on s is the set of all elements that s can be mapped to by an element of G; i.e., $orb_G(s)=\{\phi(s)|\phi \in G\}$. Then there exist $s_1, \ldots, s_n$ such that $orb_G(s_1), \ldots, orb_G(s_n)$ are disjoint and their union is S. The choice of $s_1, \ldots, s_n$ is usually not unique; however, the number of orbits of G on S, i.e., the value of n, is fixed for a given G and S. The purpose of Burnside's lemma is to calculate this number.

Theorem (Burnside's Lemma). *Let G be a finite group of permutations on a set S. Then the number of orbits of G on S is*

$$\frac{1}{|G|}\sum_{\phi\in G}|\text{fix}(\phi)|.$$

For more information about these concepts, consult an abstract algebra text such as [4] or [5].

In our problem, the group of permutations G is D_6, the dihedral group whose elements are the 12 symmetries of a regular hexagon (six rotations and six reflections). The set of all resource configurations, without removing symmetric configurations, will be the set S; recall from (1) that $|S| = 244{,}432{,}188{,}000$. For a given resource configuration s, the orbit of s is the set of all resource configurations that we can obtain by applying the symmetries in $G = D_6$ (rotations and reflections) to s.

Let's begin by calculating $\text{fix}(\phi)$ for the six rotations ϕ. When ϕ is the rotation of 0° (i.e., the identity element of D_6), ϕ fixes every element of S. Hence,

$$|\text{fix}(\phi)| = |S| = 244{,}432{,}188{,}000.$$

Furthermore, notice that if ϕ is a rotation of $(60n)^\circ$, $n = 1, \ldots, 5$, then $|\text{fix}(\phi)| = 0$; see Figure 3 and the discussion before the "A Direct Approach" section.

This leaves only the reflectional symmetries across lines of type a and b as described previously. Before proceeding, recall that there are eight available pairs of resource tiles (2 pairs of wool tiles, 2 pairs of lumber, 2 pairs of grain, 1 pair of brick, and 1 pair of ore), as well as 1 additional brick, ore, and desert.

We'll let F_a denote a reflectional symmetry (F for flip) across a line of type a. Then to fill the seven pairs of locations (marked by $x_1, \ldots, x_7$ in Figure 8a), we choose from the eight pairs of resources. We must choose the pair to exclude, place the seven pairs, and then place the remaining five resources along line a. Since there are two cases (depending on whether the excluded pair is a wool, lumber, or grain, or is a brick or ore), we have

$$\begin{aligned}|\text{fix}(F_a)| &= \binom{3}{1}\cdot\binom{7}{1,1,2,1,2}\cdot\binom{5}{2}\cdot 3! + \binom{2}{1}\cdot\binom{7}{2,2,1,2}\cdot\binom{5}{3}\cdot 2!\\ &= 252{,}000.\end{aligned}$$

Similarly, consider reflections across a line of type b. Then there are eight pairs of locations to fill with the eight pairs of resources; see Figure 8b.

The remaining three locations are filled with the leftover brick, ore, and desert. Hence,

$$|\text{fix}(F_b)| = \binom{8}{2,1,2,1,2}\cdot 3! = 30{,}240.$$

Finally, we are ready to apply Burnside's lemma. Since there are three reflectional symmetries of type a and three of type b, and Figure 7 illustrates why $F_a \cap F_b = \emptyset$, the number of distinct configurations up to symmetry is

$$\frac{1}{12}\Big(244{,}432{,}188{,}000 + 3(252{,}000) + 3(30{,}240)\Big) = 20{,}369{,}419{,}560.$$

Of course, this is the same number as we obtained using the more direct approach!

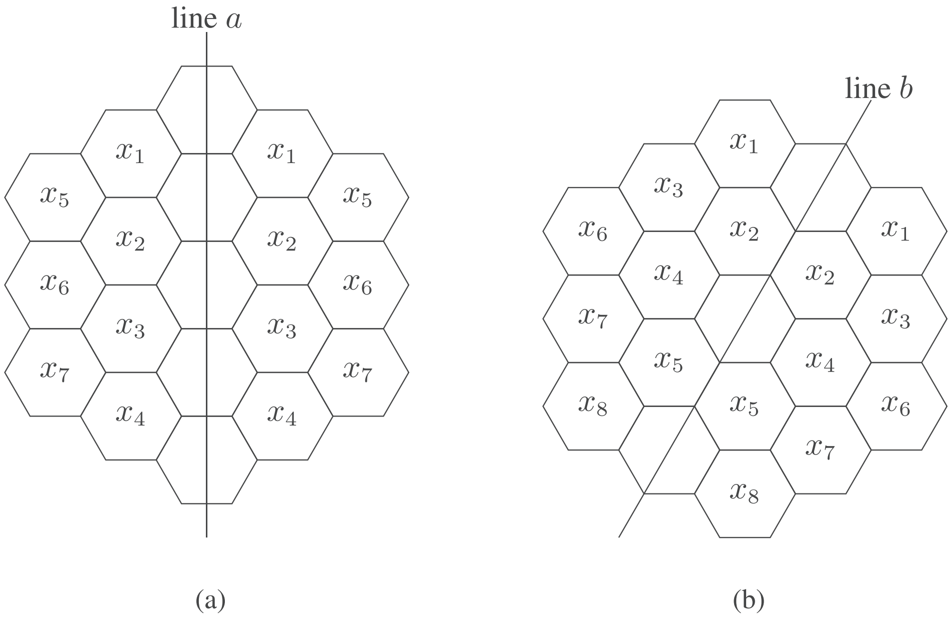


Figure 8 Configurations fixed under reflections across line *a* and line *b*, respectively.

Conclusion

The "Catanbinatorics" presented in this article provide a first insight into the combinatorial potential of this game board. The way the game itself is played provides motivation for considering additional restrictions on board configurations such as rules about resource or number adjacency, or limiting which number tokens might be placed on which resource tiles. Counting the boards within these restrictions could require still other rich combinatorial techniques. Additional counting problems may be considered for similar boards with different quantities or types of tiles. Although the number of possible boards is not actually endless, this game may provide countless opportunities for the exploration of interesting "Catanbinatorics."

Acknowledgements We would like to thank the anonymous reviewers for their helpful suggestions and Jonathon Miller for his contributions in our preliminary discussions.

REFERENCES

[1] Austin, J., Molitoris-Miller, S. (2015). *The Settlers of Catan*: Using settlement placement strategies in the probability classroom. *The College Math. J.* 46: 275–282.
[2] Catan Studio, Inc. and Catan GmbH, The official website of *Catan* (2018). http://www.catan.com
[3] Cox, C., de Silva, J., Deorsey, P., Kenter, F. H. J., Retter, T., Tobin, J. (2015). How to make the perfect fireworks display: Two strategies for *Hanabi*. *Math. Mag.* 88: 323–336.
[4] Dummit, D. S., Foote, R. M.(2004). *Abstract Algebra*. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.
[5] Gallian, J. A. (2013). *Contemporary Abstract Algebra*. 8th ed. Boston, MA: Brooks/Cole Cengage Learning.
[6] Guhe, M., Lascarides, A. (2014). Game strategies for *The Settlers of Catan*. In: *IEEE Conference on Computational Intelligence and Games* 2014. Los Alamitos, CA: IEEE Computer Society Press.
[7] Law, K. (2010). *Settlers of Catan*: Monopoly killer? *Mental Floss*. http://mentalfloss.com/article/26416/settlers-catan-monopoly-killer
[8] Mathematics Stack Exchange, *Settlers of Catan* board possibilities (2015). https://math.stackexchange.com/questions/1184670/settlers-of-catan-boards-possibilities

[9] McNary, D. (2015). *Settlers of Catan* movie TV project, *Variety*. http://variety.com/2015/film/news/settlers-of-catan-movie-tv-project-gail-katz-1201437121/

[10] Quora.com, (2013). Number of boards in *The Settlers of Catan*. https://www.quora.com/How-many-board-permutations-are-there-in-the-standard-Settlers-of-Catan-game

[11] Szita, I., Chaslot, G., Spronck, P. (2009). Monte-Carlo tree search in *Settlers of Catan*. In: Herik, H. J., Spronck, P., eds. *Advances in Computer Games*. Berlin: Springer, pp. 21–32.

[12] The Board (2006). Gameboard in *The Settlers of Catan*. https://theboard.byu.edu/questions/23546/

[13] Thomas, R., Hammond, K. (2002). Java settlers: A research environment for studying multi-agent negotiation. In: Hammond, K. J., Gil, Y., Leake, D., eds. *Proceedings of the 7th International Conference on Intelligent User Interfaces, 2002, San Francisco, California, USA, January 13–16, 2002*, Vol. 13. New York, NY: Association for Computing Machinery (ACM).

Summary. *Catan* is a dynamic property-building and trading board game in which players build a new board every time they play by arranging tiles, number tokens, and port markers. In this paper, we count the number of possible boards, consider different ways of counting the number tokens based on probability, and count the number of non-equivalent tile arrangements in two ways: one using a direct approach, the other taking advantage of more elegant techniques from abstract algebra.

JATHAN AUSTIN (MR Author ID: 1040737) is an associate professor of mathematics at Salisbury University in Maryland. He earned a B.S. in mathematics from Salisbury University, and both an M.S. in mathematics and a Ph.D. in mathematics education from the University of Delaware. He teaches a range of mathematics courses, both for future and current teachers and for traditional mathematics majors. His research interests include the mathematics of games, elementary number theory with connections to linear algebra, and the mathematical development of prospective teachers.

BRIAN G. KRONENTHAL (MR Author ID: 863868) is an associate professor of mathematics at Kutztown University of Pennsylvania. He earned his B.S. in mathematics from Lafayette College (Easton, Pennsylvania), as well as his M.S. and Ph.D. in mathematics from the University of Delaware. His favorite research problems are combinatorial, often with an algebraic flair. In addition to teaching and research, he enjoys playing ping pong, watching movies, and rooting for Philadelphia sports teams.

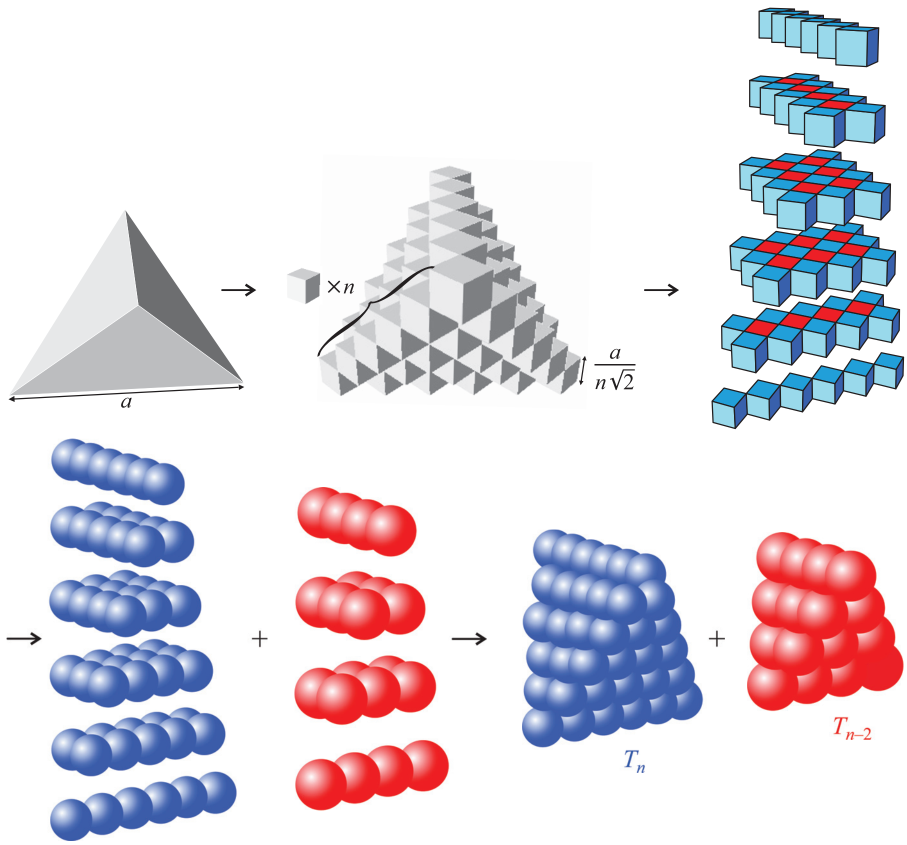
SUSANNA MOLITORIS MILLER (MR Author ID: 1284731) is an assistant professor of Mathematics Education at Kennesaw State University in GA where she teaches mathematics content courses for teachers. She earned her B.S. in mathematics from The University of Scranton, and M. S. in mathematics and Ph.D. in mathematics education from The University of Delaware. Susanna's research focuses on how students learn mathematical concepts in both formal situations, such as those involving mathematical definitions, as well as non-conventional settings, such as through games or self-directed learning experiences. In her free time, she enjoys spending time with her family, fiber arts, tea and, of course, game night with friends.

Proof Without Words: Volume of the Regular Tetrahedron via the Tetrahedral Number

ADRIAN CHUNPONG CHU
Chinese University of Hong Kong
Hong Kong, China
adrianchu001@gmail.com

I present a proof without words that a regular tetrahedron with side length a has volume $\frac{a^3}{6\sqrt{2}}$. The formula for the nth tetrahedral number, $\sum_{i=1}^{n}\sum_{j=1}^{i} j = \frac{1}{6}n(n+1)(n+2)$, is assumed. (For a proof without words of this formula, see [1].)

Proof. Recall that $T_n = 1 + 2 + \cdots + n$.



∴ vol() $= \lim_{n\to\infty}\left(\sum_{i=1}^{n} T_i + \sum_{i=1}^{n-2} T_i\right)$ vol()

$$= \lim_{n\to\infty}\left(\frac{1}{6}n(n+1)(n+2) + \frac{1}{6}(n-2)(n-1)n\right)\left(\frac{a}{\sqrt{2}n}\right)^3 = \frac{a^3}{6\sqrt{2}}.$$

Math. Mag. **92** (2019) 199–200. doi:10.1080/0025570X.2019.1586233 © Mathematical Association of America
MSC: Primary 51M05
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

REFERENCE

[1] Unal, H. (2015). Proof without words: Sums of triangular numbers. *Math. Mag.* 88: 177–178.

Summary. Using the formula for a tetrahedral number, a visual proof shows that volume of the regular tetrahedron of side length a is $\frac{a^3}{6\sqrt{2}}$.

ADRIAN CHUNPONG CHU (MR Author ID: 1240970) is a student at The Chinese University of Hong Kong. His field of interest is geometry and topology.

Math Bite: Hobbits and a Birthday-Type Problem

J.R.R. Tolkien [1, p. 35] explains how hobbits give presents to others on their birthdays:

> Hobbits give presents to other people on their own birthdays. Not very expensive ones, as a rule, and not so lavishly as on this occasion; but it was not a bad system. Actually in Hobbiton and Bywater every day in the year was somebody's birthday, so that every hobbit in those parts had a fair chance of at least one present at least once a week.

One can ask: What is the probability that some hobbit will receive a gift each day? This is the same as asking what is the probability that there are no days in which a hobbit doesn't have a birthday. This is reminiscent of the birthday problem. If there are $n \geq 365$ hobbits, then this probability is given by the following expression:

$$\frac{\displaystyle\sum_{\substack{k_i > 0 \\ k_1+\cdots+k_{365}=n}} \binom{n}{k_1, k_2, \ldots, k_{365}}}{\displaystyle\sum_{\substack{k_i \geq 0 \\ k_1+\cdots+k_{365}=n}} \binom{n}{k_1, k_2, \ldots, k_{365}}} = \frac{\displaystyle\sum_{\substack{k_i > 0 \\ k_1+\cdots+k_{365}=n}} \binom{n}{k_1, k_2, \ldots, k_{365}}}{365^n},$$

where k_i is the number of hobbits that have a birthday on day i. The denominator is equal to 365^n by an application of the multinomial theorem,

$$(x_1 + x_2 + \cdots + x_\ell)^n = \sum_{\substack{k_i \geq 0 \\ k_1+\cdots+k_\ell=n}} \binom{n}{k_1, \ldots, k_\ell} x^{k_1} \cdots x^{k_\ell},$$

by letting $x_i = 1$, for all i, and $\ell = 365$.

REFERENCE

[1] Tolkien, J. R. R. (1965) *The Fellowship of the Ring: Being the First Part of The Lord of the Rings*, 2nd ed., Boston: Houghton Mifflin Company.

—Submitted by Stanley R. Huddy (srh@fdu.edu)
Fairleigh Dickinson University

Wendy Carlos's Xenharmonic Keyboard

JORDAN SCHETTLER
Department of Mathematics and Statistics
San José State University, San Jose, CA
jordan.schettler@sjsu.edu

Wendy Carlos is an American composer and electronic musician; as a musician and composer myself, she is one of my favorite musical pioneers. She helped to popularize and improve the Moog synthesizer with the Grammy winning album *Switched on Bach* in 1968. I recall fondly when my grandfather gave me his copy of that record on vinyl. With an orchestrated flurry of futuristic sounding beeps, I didn't know at first if I was playing it at the right speed! Some of Carlos's other best known works are the motion picture soundtracks for Stanley Kubrick's *A Clockwork Orange* in 1971, *The Shining* in 1980, and Walt Disney's *Tron* in 1982.

Here we are interested in Carlos's 1986 original work *Beauty in the Beast* because of the experimentation with various so-called *xenharmonic scales*, i.e., scales other than the familiar 12-tone scale on a standard keyboard. The standard keyboard is equally spaced with 100 cents between consecutive notes. In contrast, the title track on *Beauty in the Beast* uses a 9-tone and 11-tone scale with $\alpha = 77.995...$ and $\beta = 63.814...$ cents, respectively, between consecutive notes. The α and β scales are both "contained" in a scale with $\gamma = 35.097...$ cents between consecutive notes. In her article "Tuning: At the Crossroads" [3], Wendy Carlos explains how she derived these scales by experimentation. In contrast, we will derive these scales using the theory of continued fractions in the spirit of Dunne and McConnell's article "Pianos and Continued Fractions" [4] or Section 6.6 in Dave Benson's *Music: A Mathematical Offering* [2]. This method will further demonstrate how the containment of the α and β scales inside the γ scale is analogous to the way in which the pentatonic scale (5 black keys) and heptatonic scale (7 white keys) fit into the familiar 12-tone scale found on a standard keyboard. Finally, we will illustrate this analysis by drawing Carlos' γ scale keyboard. The interspersing of black and white keys comes from the topological interpretation of notes as points on the unit circle in the complex plane.

The topological group of notes

Given an ideal taut string with fixed endpoints (think of strings on a piano, violin, or guitar), vibrations lead to periodic compression waves in air which we perceive as sound. We define the *fundamental period* T to be the number of seconds in one minimal cycle of these periodic vibrations (i.e., T is the shortest length of time such that the vibrations repeat themselves every T seconds). The *fundamental frequency* $f = 1/T$ is the number of minimal cycles per second. The units of f here are measured in Hertz:

$$1\,\text{Hz} = 1\,\frac{1}{\text{sec}}.$$

We will not write the units Hz, and we will simply regard f as a number in the interval of positive real numbers $(0, \infty)$.

We perceive differences in frequency as differences in pitch. For example, the frequency of *Middle C* on a piano is around 261.6, which we recognize as a lower pitch

Math. Mag. **92** (2019) 201–212. doi:10.1080/0025570X.2019.1569938
MSC: Primary 00A65, Secondary 11A55
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/umma.

than that of *Concert A* with a frequency of 440. Various physical properties of the string govern its fundamental frequency, and our ideal string may vibrate at multiple frequencies simultaneously not just the fundamental frequency. These other frequencies, which are positive integer multiples of the fundamental frequency f, are called *harmonics*: $f, 2f, 3f, 4f, \ldots$.

The pitches corresponding to f and $2f$ really do sound alike, even though $2f$ corresponds to a higher pitch. In fact, they sound so similar that we consider them to represent the same note. For example, if $f = 220$, then $2f = 440$ is our Concert A, so this f should also represent an A note, just with a lower pitch. In general, we say $2f$ is one *octave* higher than f. Likewise, we say the frequency $2^{-1}f$ is one octave lower than f. In this way, all the frequencies of the form $2^n f$ for $n \in \mathbb{Z}$ determine octaves of the same note. There is a natural equivalence relation on frequencies $f, g \in (0, \infty)$:

$$f \sim g \Leftrightarrow f = 2^n g \text{ for some integer } n.$$

We now define a *note* to be the equivalence class of a frequency

$$[f] := \{g \in (0, \infty) \colon f \sim g\}.$$

Thus [440] is the A note in the sense that this class is exactly the set of frequencies which represent an A note. We define the *note space*

$$N_2 := \{[f] \colon f \in (0, \infty)\},$$

where the subscript 2 indicates that two frequencies are related precisely when their quotient is an integer power of 2. There are infinitely many notes in N_2 of course, since every note is of the form $[f]$ for some unique $f \in [1, 2)$.

Our note space N_2 has a mathematical structure beyond just being a set of tones. The interval of positive real numbers $(0, \infty)$ is a group under multiplication. The set $\langle 2 \rangle$ of integer powers of 2 is a subgroup of $(0, \infty)$, so our note space is a quotient group $N_2 \cong (0, \infty)/\langle 2 \rangle$ with the product of notes given by $[f] \cdot [g] = [f \cdot g]$.

We also want to make sense of the distance between two notes. Since N_2 is in one-to-one correspondence with the interval $[1, 2)$ where $[1] = [2]$, we should think of N_2 as a line segment with its endpoints identified. In other words, N_2 should be a circle. Indeed, there is a way of making this precise. The unit circle $\mathbb{S}^1$ in the complex plane $\mathbb{C}$ is a nice model of a circle since it is naturally a group under the usual multiplication of complex numbers and it can be described very simply as the set of complex numbers of the form $e^{i\theta}$ with $\theta \in [0, 2\pi)$. These $e^{i\theta}$ are complex numbers with distance 1 from the origin where the angle θ is taken with respect to the positive real axis. Actually, we can take this θ to be any real number since the function e^{ix} is periodic in x with period 2π. If $[f] = [g]$, then $f = 2^n g$ for some $n \in \mathbb{Z}$, so the properties of logarithms imply

$$\log_2(f) = \log_2(2^n g) = n + \log_2(g),$$

and the periodicity of the exponential further gives

$$e^{2\pi i \log_2(f)} = e^{i(2\pi n + 2\pi \log_2(g))} = e^{2\pi i \log_2(g)}.$$

Therefore we have a well-defined map from our note space to the unit circle

$$N_2 \longrightarrow \mathbb{S}^1 : [f] \mapsto e^{2\pi i \log_2(f)}.$$

Moreover, this map is an isomorphism, meaning that it represents a one-to-one correspondence which additionally preserves the group structures of N_2 and $\mathbb{S}^1$:

$$e^{2\pi i \log_2(f \cdot g)} = e^{2\pi i \log_2(f)} e^{2\pi i \log_2(g)}.$$

Thus we may regard multiplication by the note $[f]$ as a counterclockwise rotation of the circle through an angle of $2\pi \log_2(f)$ radians.

There is an obvious notion of distance in $\mathbb{S}^1$ obtained by simply considering the distance of points in the complex plane. We wish to consider a different (though topologically equivalent) measure of the distance between two points, namely we define the distance between P and Q on $\mathbb{S}^1$ to be the radian measure θ of the non-obtuse angle formed by the segments joining P and Q with the origin. Without loss of generality, we have $P = e^{i\theta_1}$ and $Q = e^{i\theta_2}$ with $0 \le \theta_1 \le \theta_2 < 2\pi$, so the distance between P and Q is the minimum of $\theta_2 - \theta_1$ and $2\pi + \theta_1 - \theta_2$ as in Figure 1. This distance

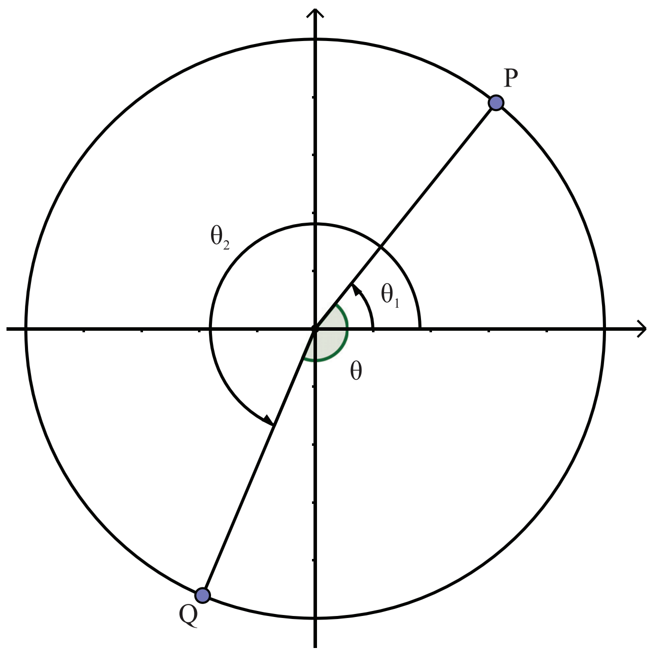


Figure 1 Distance between P and Q is θ.

turns $\mathbb{S}^1$ into a topological group, which means that multiplication and inversion are continuous operations: given convergent sequences on $\mathbb{S}^1$, say $\lim_{n\to\infty} P_n = P$ and $\lim_{n\to\infty} Q_n = Q$, then

$$\lim_{n\to\infty} P_n \cdot Q_n = P \cdot Q \quad \text{and} \quad \lim_{n\to\infty} P_n^{-1} = P^{-1},$$

where $P_n^{-1} = 1/P_n$ and $P^{-1} = 1/P$ are reciprocals. We can give our note space this inherited distance, and then N_2 becomes a topological group that is isomorphic to $\mathbb{S}^1$.

The perfect fifth

Let us fix a frequency $f \in (0, \infty)$. As frequencies range over the octave interval $[f, 2f)$, the corresponding points on $\mathbb{S}^1$ represent one full trip around the circle, so every note has a representative in this octave. What about the non-octave harmonics? For example, neither $3f$ nor $5f$ are octaves of f, but both are natural frequencies to consider along with f. We have $[3f] = [(3/2)f]$ and $[5f] = [(5/4)f]$ with $3/2, 5/4 \in [1, 2)$. Thus $(3/2)f$ and $(5/4)f$ are frequencies lying strictly within the octave range $[f, 2f)$, and both sound pleasant when played with f. We say $(3/2)f$ is one *perfect fifth* above f and $(5/4)f$ is one *major third* above f, so we get two "new" notes. We could also consider the frequencies $(4/3)f$, $(8/5)f$ which are a perfect fifth

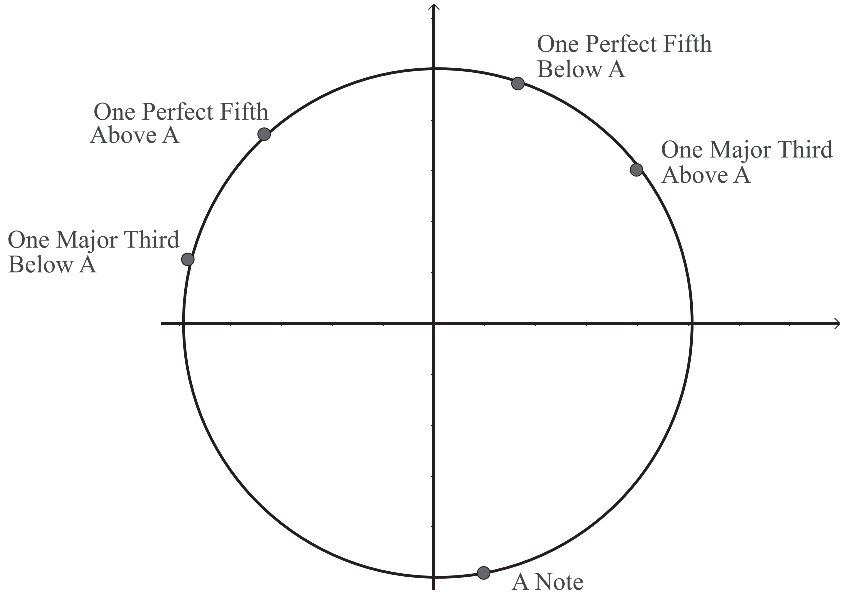


Figure 2 Perfect fifth and major third.

and major third below $2f$. Again we get two more "new" notes. We have plotted these notes on the unit circle in Figure 2 for the case $[f] = [440]$ is A.

What about higher harmonics? For example, $[6f] = [(3/2)f]$, $[10f] = [5f] = [(5/4)f]$, etc., are not new notes, but $[7f] = [(7/4)f]$, $[11f] = [(11/8)f]$, etc., are new notes. Of course, there are infinitely many notes coming from harmonics of f which are not products of previous notes since there are infinitely many primes. These less simple ratios $(7/4)f$, $(11/8)f, \ldots$, are considered to sound less consonant when played with f. For this reason, we first consider the problem of trying to construct a scale using only octaves and perfect fifths. In other words, we want a musical scale (collection of frequencies) such that if f is in the scale, then so are the octave $2f$ and the perfect fifth $(3/2)f$ because these are the two most pleasant sounding ratios.

One problem with this idea is that the number of new notes in N_2 we will obtain is infinite! To see why this is so, let's begin with the frequency of $f = 1$; we may do this since we could later rotate the circle $\mathbb{S}^1$ so that our scale begins with any frequency we like. Taking octaves gives a sequence $1, 2, 4, 8, \ldots$, which all represent the same note, but taking perfect fifths gives a sequence

$$(3/2)^0 = 1, \qquad (3/2)^1 = 3/2, \qquad (3/2)^2 = 9/4,$$
$$(3/2)^3 = 27/8, \qquad (3/2)^4 = 81/16, \qquad \ldots,$$

which we will now show all represent distinct notes. Two of these frequencies $(3/2)^m$ and $(3/2)^n$ for nonnegative integers $m < n$ represent the same note if and only if the corresponding points on the circle are equal, i.e., $e^{2\pi i m \log_2(3/2)} = e^{2\pi i n \log_2(3/2)}$. Equivalently, $(n-m)\log_2(3/2) = k$ for some positive integer k, but this leads to $3^{n-m} = 2^{k+n-m}$, contradicting the fundamental theorem of arithmetic. This proves that $\log_2(3/2)$ is an irrational number, which, in turn, is equivalent to the fact that the subgroup $\langle[3/2]\rangle$ of the note space N_2 generated by $[3/2]$ is infinite. More generally, $\langle[f]\rangle$ is infinite if and only if $\log_2(f)$ is irrational.

We know that we cannot stack perfect fifths indefinitely, but when should we stop?

Again beginning with a frequency $f = 1$, we get notes

$$[1], [3/2], [9/8], [27/16], [81/64], \ldots.$$

where we have chosen frequency representatives to lie in the interval $[1, 2)$. The first five notes with these representatives satisfy the inequalities

$$1 < 9/8 < 81/64 < 3/2 < 27/16 < 2 \tag{1}$$

and the distances (i.e., angle measures) between the consecutive notes on $\mathbb{S}^1$ range from about 1.06767 to 1.54.

Something interesting happens with the representative of the sixth note:

$$[(3/2)^5] = [243/128] \text{ with } 27/16 < 243/128 < 2$$

This means the sixth note is closer to the first note than any of the previous four. The distance on $\mathbb{S}^1$ between the first and sixth note is around 0.472417 which is significantly less than the gaps between the first five notes. The sixth note does not represent a great approximation to the first note because humans can distinguish between two notes which are a distance of more than 0.026 apart (see [6]), although the ability to discern different pitches varies from person to person, and also depends on both amplitude and timbre.

We can proceed in two ways. First, we could just pretend the frequency $243/128 = 1.8984375$ is 2. Then we just repeat the arrangement of the frequencies from equation (1) in every octave to create our scale:

$$\cdots < 3/4 < 27/32 < 1 < \cdots < 2 < 9/4 < 81/32 < \cdots.$$

Although we have true perfect fifths above some frequencies like 1 (since $3/2$ is in our scale), we do not for others like $81/64$ (since $243/128 \neq 2$ is not in our scale). For this reason, we cannot always translate a piece of music written in this scale from one key to another.

Alternatively, we could distribute the error between the sixth note and the first among all the five prior notes. One way to do this is to make the gaps between consecutive notes equal. In other words, each note on $\mathbb{S}^1$ would be a distance of $2\pi/5 \approx 1.256637$ from its neighboring notes. Five equally spaced notes are of the form $e^{2\pi i(k/5)} \leftrightarrow [2^{k/5}]$ for $k = 0, 1, 2, 3, 4$. Now the sixth note $e^{2\pi i(5/5)} = 1 = e^{2\pi i(0/5)}$ is the same as the first, but the approximation to the perfect fifth $[3/2] \overset{?}{\approx} [2^{3/5}]$ is not great since the distance on $\mathbb{S}^1$ is

$$2\pi |\log_2(3/2) - 3/5| \approx 0.0944834,$$

which is well above the lower bound of 0.026 where humans can distinguish notes. This equal distribution of notes, called an *equal temperament*, has a major added benefit: we can transpose any piece of music in any way we like. This means that a piano score written in one key can easily be shifted to another key while preserving all the relative distances of notes in the work. As long as $f = 1$ is in our scale (by appropriately scaling our units of frequency), this transposition amounts to multiplying every frequency in the score by a fixed frequency in our scale. In algebraic terms, this means that our set of five equally spaced notes forms a finite subgroup of the note space N_2. In particular, the notes are invertible as well: $e^{2\pi i(1/5)}e^{2\pi i(4/5)} = 1 = e^{2\pi i(2/5)}e^{2\pi i(3/5)}$, so we can go up or down any number of approximate perfect fifths we like. Scales with five notes per octave are called *pentatonic*. We have just described two pentatonic scales, one equally tempered and the other not. The non-equally tempered scale had some true perfect fifths but was not transposable, while the equally tempered scale was transposable but did not have a great approximation to the perfect fifth. One way to get the best of both worlds is to increase the number of notes in our equally tempered scale to get a better approximation to the perfect fifth.

If we continue stacking true perfect fifths beyond the sixth, then the next time we get closer to where we started is on the thirteenth note, which has a distance on $\mathbb{S}^1$ of around 0.12284 from the first note. Now we distribute this error among all the twelve prior notes by considering an equally tempered scale with 12 notes. Here our notes on $\mathbb{S}^1$ are all of the form $e^{2\pi i(k/12)}$ for $k = 0, 1, \ldots, 11$, and our approximation to the perfect fifth becomes

$$2\pi\,|\log_2(3/2) - 7/12| \approx 0.01023636,$$

which is below the lower bound 0.026 at which humans can distinguish notes. In other words, we would perceive the pitches $\log_2(3/2)$ and $7/12$ to be the same, so this is an excellent approximation. In Figure 3, we have plotted the 12 notes from this

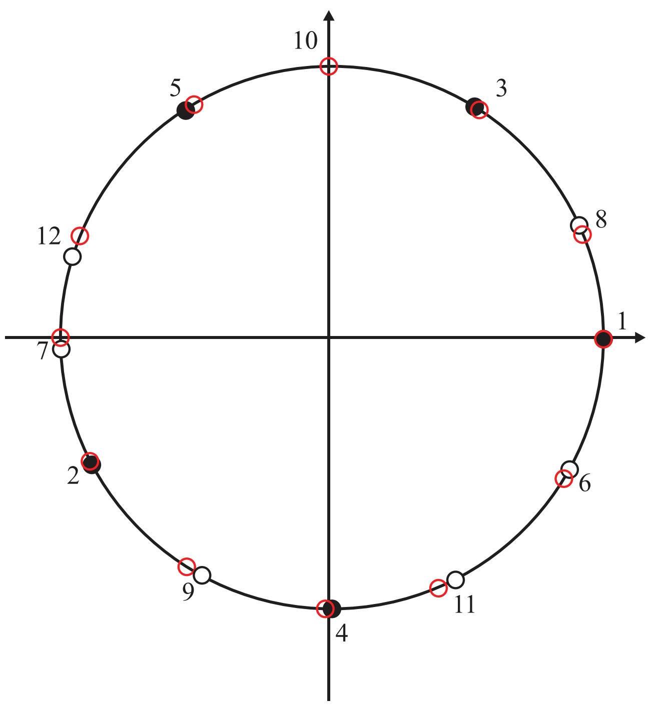


Figure 3 Stacked perfect fifths (black and white dots) and equally spaced approximations (red circles).

equally tempered scale alongside the stacked true perfect fifths that each note represents. Notice how close note number 2 (true perfect fifth) is to its approximation. This first five stacked perfect fifths are colored black, and the last seven are colored white. Coloring the notes this way allows us to see that our equally tempered 12-note scale contains an approximation to the non-equally tempered pentatonic scale we constructed earlier. We also get a non-equally tempered seven note, or *heptatonic*, scale from the white notes. If we cut and straighten our note space/circle $N_2 \cong \mathbb{S}^1$, then this arrangement of black and white notes corresponds precisely to the keys in one octave of a standard keyboard as seen in Figure 4. Since the 12-note equal temperament

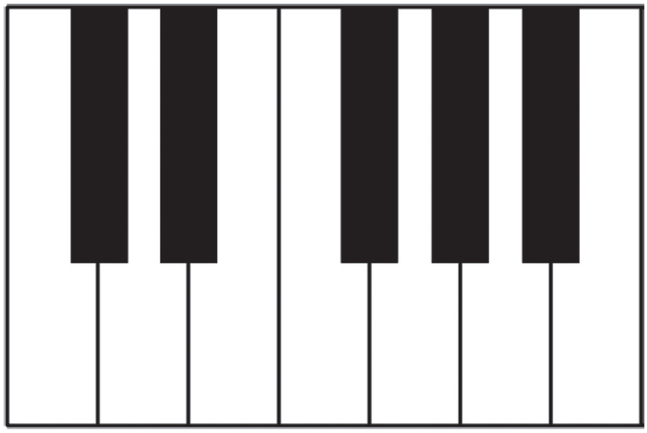


Figure 4 One octave on a standard keyboard.

scale is so prevalent, the distance between notes or frequencies is often measured in *cents*:

$$1200 \text{ cents} = 2\pi \text{ radians},$$

so consecutive frequencies in this scale are exactly 100 cents apart. In these units, humans cannot distinguish frequencies less than 5 cents apart, and the 12-note equally tempered scale has a decent, but not great, approximation to the major third (see note number 5 in Figure 3):

$$2\pi \,|\log_2(5/4) - 4/12| \approx 0.07166 \approx 13.686 \text{ cents}. \quad (2)$$

What if we wanted better approximations to the perfect fifth and major third? We could add more notes, but the next time our perfect fifth stacking gets closer to the first note than all of the previous ones is on the forty-second note. This would give us 41 notes in each octave! Where do these numbers of notes 5, 12, 41 come from? Could we have found them in a simpler way?

Continued fractions

We want an algorithm to determine the numbers n of stacked perfect fifths such that $[(3/2)^n]$ is closer to $[1]$ than any previous note. For a positive integer b, the distance between $[(3/2)^b]$ and $[1]$ is

$$2\pi |b \log_2(3/2) - a|, \quad (3)$$

where a is the unique positive integer which ensures this quantity is less than π, i.e.,

$$|b \log_2(3/2) - a| < 1/2,$$

so such an a minimizes the expression (3). Here a/b is a rational approximation to the irrational number $\log_2(3/2)$ which becomes arbitrarily better by choosing a sufficiently large denominator b or, equivalently, choosing more notes per octave. Ideally, we would like to do this in an efficient way, which means getting a good approximation while using the minimal number of notes. Equivalently, we want to find positive integers n and m such that

$$|b \log_2(3/2) - a| > |n \log_2(3/2) - m|,$$

whenever $0 < b < n$ and a is any integer. In this way, m/n will be the "best" rational approximation to $\log_2(3/2)$ having denominator less than or equal to n. Such best approximations are characterized as being convergents of continued fractions. As we will see, these continued fractions and their convergents can be successively computed in a systematic way.

For a real number a_0 and positive real numbers $a_1, \ldots, a_k$, we define the *finite continued fraction* by

$$[a_0; a_1, a_2, \ldots, a_k] := a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots \cfrac{1}{a_k}}}.$$

We define an *infinite continued fraction* to be the limiting value of finite continued factions, where

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}} := \lim_{k\to\infty} [a_0; a_1, a_2, \ldots, a_k].$$

It turns out that this limit exists if and only if the infinite sum $\sum_{i=0}^{\infty} a_i$ diverges (see [5]). We assume now that all the a_i are integers, so the infinite continued fraction always exists in this case. In fact, much more is true.

Theorem 1. *Let α be an irrational real number. There is a unique infinite* continued fraction expansion

$$\alpha = [a_0; a_1, a_2, \ldots],$$

such that a_0 is an integer and $a_1, a_2, \ldots$ are positive integers.

The uniqueness claim in this theorem is not difficult to show. Hence the real content of the theorem is proving that this sequence of a_i's so-defined actually converges to the irrational α you started with. That is indeed the case (see [5]), so these digits a_i in the continued fraction expansion are somewhat like decimals in a decimal expansion, in the sense that computing more digits gives you a progressively better rational approximation.

Here are a few examples of continued fraction expansions of irrational real numbers:

$$\begin{aligned} \sqrt{11} &= [3; 3, 6, 3, 6, 3, 6, 3, 6, 3, 6, 3, \ldots], \\ e &= [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots], \text{ and} \\ \pi &= [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, \ldots]. \end{aligned}$$

The suggested patterns in the expansions of $\sqrt{11}$ and e continue forever, but there is no known pattern in the continued fraction expansion of π.

We define the *kth convergent* of a real irrational

$$\alpha = [a_0; a_1, a_2, \ldots] \tag{4}$$

to be the value of the finite continued fraction $[a_0; a_1, a_2, \ldots, a_k]$. The 0th convergent is $[a_0] = a_0/1$, the first convergent is $[a_0; a_1] = (a_1a_0 + 1)/a_1$, and the second convergent is

$$[a_0; a_1, a_2] = \frac{a_2(a_1a_0 + 1) + a_0}{a_2a_1 + 1}.$$

In general, one can prove that the kth convergent equals p_k/q_k where

$$p_k = a_k p_{k-1} + p_{k-2} \text{ and } q_k = a_k q_{k-1} + q_{k-2}, \tag{5}$$

with initial conditions

$$p_0 = a_0, \qquad p_1 = a_1 a_0 + 1, \qquad q_0 = 1, \qquad q_1 = a_1. \tag{6}$$

These numerators p_k and denominators q_k of convergents give us precisely the type of best approximations we are looking for.

Theorem 2. *Given an irrational real number α, the convergents p_k/q_k (as determined by equations (4)–(6)) are in lowest terms and represent best approximations in the following sense.*

For any integers a, b with $0 < b < q_k$, we have

$$|b\alpha - a| > |q_k\alpha - p_k|.$$

Conversely, if m, n are integers with $n > 0$ such that

$$|b\alpha - a| > |n\alpha - m|$$

whenever a, b are integers with $0 < b < n$, then $m = p_k$ and $n = q_k$ for some k.

This theorem (proved in [5]) tells us that the number n of notes we should divide the octave into will be the denominator q_k of some convergent in the continued fraction expansion

$$\log_2(3/2) = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{3 + \cdots}}}}}.$$

In general, the convergents of α satisfy the inequalities

$$\frac{p_2}{q_2} < \frac{p_4}{q_4} < \frac{p_6}{q_6} \cdots < \alpha < \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1},$$

and for $\alpha = \log_2(3/2)$ we have

$$\frac{1}{2} < \frac{7}{12} < \frac{31}{53} < \cdots < \log_2(3/2) < \cdots < \frac{24}{41} < \frac{3}{5} < \frac{1}{1}.$$

Here we see the denominators 5, 12, 41, 53, ..., counting the numbers of notes needed per octave in order to get the successively next best approximation to the perfect fifth using equal temperament. The numerators tell us which note in the equally tempered scale approximates the perfect fifth. For example, the convergent 7/12 tells us that the approximation to a perfect fifth in the equally tempered 12-note scale occurs 700 cents above a given frequency. Likewise, the convergent 24/41 tells us that the approximation to a perfect fifth in the equally tempered 41 note scale occurs $1200 \cdot (24/41) = 702.\overline{43902}$ cents above a given frequency.

What about an equally tempered heptatonic scale? Our equally tempered 12-note scale contains a non-equally tempered heptatonic scale (7 white keys), but 7 is not the denominator of a convergent for $\log_2(3/2)$, so there is no reason to expect an equally tempered heptatonic scale to be desirable. However, it is a somewhat desirable scale since 7 is a denominator in one of the "intermediate fractions"

$$\frac{0 \cdot 3 + 1}{0 \cdot 5 + 2} = \frac{1}{2} < \frac{1 \cdot 3 + 1}{1 \cdot 5 + 2} = \frac{4}{7} < \frac{2 \cdot 3 + 1}{2 \cdot 5 + 2} = \frac{7}{12}.$$

In general, an *intermediate fraction* is of the form

$$\frac{tp_{k-1} + p_{k-2}}{tp_{k-1} + p_{k-2}} \qquad \text{where } 0 \leq t \leq a_k.$$

These are also best approximations, but in a weaker sense than Theorem 2: if r, s are integers with $s > 0$ such that $|\alpha - a/b| > |\alpha - r/s|$ whenever a, b are integers with $0 < b < s$, then r/s is an intermediate fraction. For this to be true as stated we need to include the index $k = -1$: $p_{-1} = 1$, $q_{-1} = 0$.

We should pay attention to the fact that our decent major third approximation in equation (2) was essentially coincidental. We focused solely on approximations to the perfect fifth, but there are techniques involving so-called ternary continued fractions which give constructions for finding simultaneous rational approximations to two or more irrational real numbers. See [**1**] for details on the applications of ternary continued fractions to music.

Xenharmonic scales

Why did we have to start with an octave interval $[1, 2)$? In other words, why must we identify the frequencies f and $2f$ as representing the same note? We could choose, for instance, to identify f with higher harmonics like $3f$ or $5f$, or we could work with intervals smaller than one octave. Scales which are not based on the octave interval are xenharmonic (as defined at the beginning of this article). For example, in the article [**3**], Wendy Carlos started instead with the perfect fifth interval $[1, 3/2)$, and then considered dividing this into equal pieces. Carlos' approach for dividing this perfect fifth interval was largely experimental. She picked some notes (including the major third) that she wanted to be well approximated, then gradually incremented step sizes and computed (minus) the total squared deviations between the ideal frequencies and the approximations thereof. She found the following desirable divisions in Table 1.

Step Sizes	Number of Notes
$\alpha = 77.995...$ cents	9
$\beta = 63.814...$ cents	11
$\gamma = 35.097...$ cents	20

TABLE 1: Desirable divisions.

All three of these scales fit beautifully into our continued fraction apparatus if we simply replace the octave and perfect fifth with the prefect fifth and the major third, respectively. To see why this is so, we note that our equivalence relation now becomes

$$f \sim g \Leftrightarrow f = (3/2)^n g \text{ for some integer } n,$$

so we get a new note space $N_{3/2}$ consisting of these equivalence classes $[f]$. Again, the note space is isomorphic to the unit circle $\mathbb{S}^1$ as topological groups:

$$N_{3/2} \longrightarrow \mathbb{S}^1 : [f] \mapsto e^{2\pi i \log_{3/2}(f)}.$$

Now every note $[f]$ has a unique representative in $[1, 3/2)$. The most significant frequency lying strictly within the interval $[1, 3/2)$ is $5/4$, coming from the major third. To use the continued fraction approach, we must consider rational approximations of $\log_{3/2}(5/4)$ coming from a continued fraction expansion:

$$\log_{3/2}(5/4) = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{2 + \cfrac{1}{6 + \cdots}}}}}.$$

In this case, we have a chain of inequalities of convergents

$$\frac{1}{2} < \frac{11}{20} < \frac{82}{149} < \cdots < \log_{3/2}(5/4) < \cdots < \frac{71}{129} < \frac{5}{9} < \frac{1}{1}.$$

Immediately, we see Carlos' α and γ scales coming from the denominators 9 and 20, respectively. The β scale can be seen in the denominator 11 of an intermediate fraction:

$$\frac{0 \cdot 5 + 1}{0 \cdot 9 + 2} = \frac{1}{2} < \frac{1 \cdot 5 + 1}{1 \cdot 9 + 2} = \frac{6}{11} < \frac{2 \cdot 5 + 1}{2 \cdot 9 + 2} = \frac{11}{20}.$$

Therefore this $9 + 11 = 20$ division of the perfect fifth is in striking analogy to the $5 + 7 = 12$ division of the octave. It is worth mentioning that in any of the equally-tempered worlds α, β, or γ, the perfect fifths are true perfect fifths, we get good (and sometimes great) approximations to the major third, but the octave is no longer a priority! For instance, the error in the approximation to the octave in the gamma scale is given by ≈ 6.6765 cents, which is not very good at all considering we are using around 34 notes per octave.

If we follow the approach from The Perfect Fifth section above, we can stack major thirds around a perfect fifth circle. In analogy with the equally tempered 12-note scale, we color the first 9 notes black, and the remaining 11 notes white.

The result is seen in Figure 5. Again, we can cut and straighten our note space/circle $N_{3/2} \cong \mathbb{S}^1$, and this arrangement of black and white notes corresponds precisely to the keys in one perfect fifth of Carlos' γ scale keyboard, as seen in Figure 6. This xenharmonic keyboard has twenty keys per perfect fifth, with 9 black keys corresponding to a non-equally tempered version of the α-scale and 11 white keys corresponding to a non-equally tempered version of the β-scale. The cents-per-note can be calculated as

$$\frac{1200 \log_2(3/2)}{20} = 35.0977500432\ldots.$$

As a musician, composer, and mathematician, I would love to see this γ-keyboard physically constructed and played. That would be a wonderful tribute to mathematics, music, and creativity.

REFERENCES

[1] Barbour, J. M. (1948). Music and ternary continued fractions. *Amer. Math. Monthly.* 55(9): 545–555.

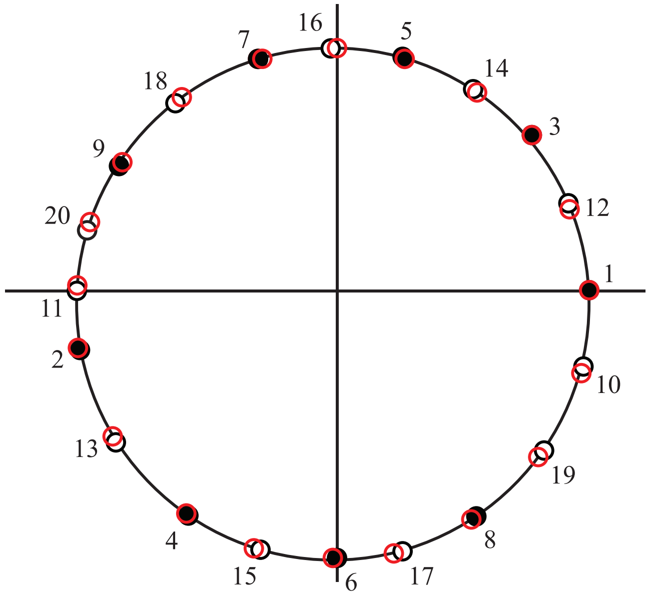


Figure 5 Stacked major thirds (black and white dots) and equally spaced approximations (red circles).

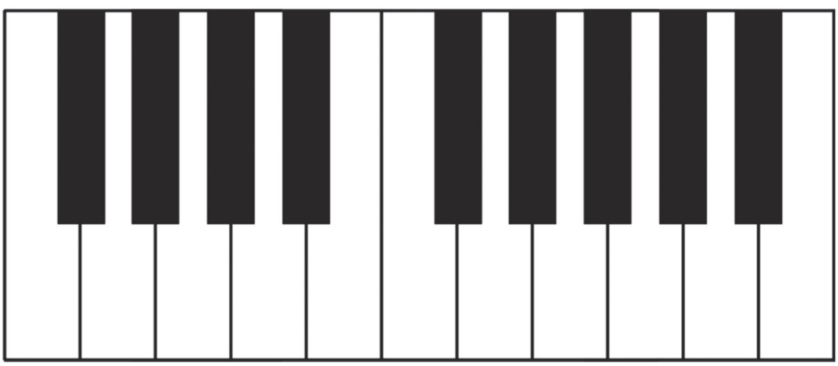


Figure 6 One perfect fifth on Wendy Carlos's keyboard.

[2] Benson, D. (2008). *Music: A mathematical offering*. Available at http://www.maths.abdn.ac.uk/, pp. 517+xiv.
[3] Carlos, W. (1987). Tuning: At the crossroads. *Comput. Music J.* 11(1): 29–43.
[4] Dunne, E., McConnell, M. (1999). Pianos and continued fractions. *Math. Mag.* 72(2): 104–115.
[5] Khinchin, A. Y. (1992). *Continued Fractions*. 3rd ed. Mineola, NY: Dover.
[6] Loeffler, D. B. (2006). Instrument timbres and pitch estimation in polyphonic music. Master's thesis, Georgia Tech, April 2006, Department of Electrical and Computer Engineering.

Summary. Wendy Carlos is an American composer and electronic musician who constructed three closely related musical scales she called α, β, and γ. These scales are xenharmonic, i.e., unrelated to the familiar 12-tone scale. However, the 5 black and 7 white keys contained in one octave on a standard keyboard have a strong analogy with representations of the α and β scales contained in one perfect fifth on the γ scale. We explore this analogy using idea from algebra and number theory, and find that Carlos's three scales arise naturally from the theory of continued fractions. Moreover, an infinite family of similarly "nice" musical scales arises from the same construction.

JORDAN SCHETTLER (MR Author ID: 1053906) received his Ph.D. in mathematics from the University of Arizona under the direction of William McCallum in 2012. He is now an assistant professor of mathematics at San José State University. His research interests include number theory and connections of mathematics to art, music, and politics.

The Chicken Braess Paradox

KIMMO ERIKSSON
Mälardalen University
Västerås, Sweden
kimmo.eriksson@mdh.se

JONAS ELIASSON
Stockholm Transport Administration
Stockholm, Sweden
jonas.eliasson@stockholm.se

Consider a small region consisting of A-town and B-town. The two towns are connected by two parallel roads that pass through either C-junction or D-junction. These roads are old and bumpy and can be traveled only at a very modest speed that does not depend on the traffic volume. To drive from A-town to either junction takes one hour and then it is another hour's drive to reach B-town. There is an even older crossroad between C-junction and D-junction, but it is in very poor condition and nobody uses it. See Figure 1 (Left).

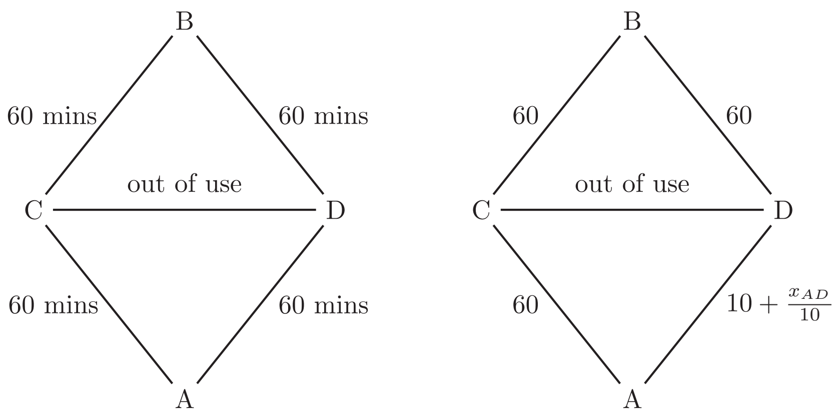


Figure 1 (Left) The A-town-B-town region's original road system with travel times for each road section. (Right) Travel times after improvement of road section AD.

Despite the bumpy roads, 200 cars drive in each direction between A-town and B-town during the morning rush. Thus, there are 200 *A-drivers* going from A-town to B-town and 200 *B-drivers* going in the opposite direction. The council has decided to achieve a quicker commute by improving the roads. One council member, Mrs. Jones, happens to be a mathematics graduate. She tells the chairman she could try to calculate which sections of the road system to improve. The chairman brushes her offer aside: "Mathematics schmathematics, who needs it? It is a no-brainer that *any* road improvement makes travel quicker!" The council decides to start the project by improving the section between A-town and D-junction.

Math. Mag. **92** (2019) 213–221. doi:10.1080/0025570X.2019.1571375
MSC: Primary 91A43

At the council meeting following the completion of the first section, the chairman proudly announces that the time to travel between A-town and B-town has now decreased by 30 minutes, from two hours to an hour and a half, and the improvement project will therefore continue with another section. Mrs. Jones tries to offer a suggestion but the chairman interrupts her: "No suggestions from you are required, Mrs. Jones. I have already told the workers to continue with the section between B-town and C-junction."

The second section is completed. At the next council meeting, the chairman is a little testy. The time to travel between A-town and B-town has decreased further, but only by a meagre 10 minutes. Mrs. Jones is about to explain why. However, she is silenced by an angry look from the chairman who has an announcement to make: "I have come to the insight that we need to tie the two improved sections together. I have ordered an improvement of the crossroad between C-junction and D-junction next." Because of budget constraints the council can only make this improved crossroad a single lane road. The chairman happens to be an A-driver and he has decided that the crossroad will only be open one-way, for traffic going from D-junction to C-junction, thus enabling A-drivers to use all three improved sections.

When the council meets after the completion of the third section, the chairman is a broken man. To his utter disbelief the time to travel from A-town to B-town has now *increased* by 10 minutes. He has checked that the amount of traffic has not increased; the same number of cars travel from A-town to B-town as always. He has also made sure that local radio continually gives accurate information about the traffic situation so that drivers can make optimal decisions—to no avail. Somehow his improvement made travel slower. How could this be?

When Mrs. Jones raises her hand to offer an explanation, the poor chairman is too deflated to stop her. Mrs. Jones points out that the improved roads are certainly much faster than the old roads, but that the very highest speed can only be achieved if you are alone on the road; the more traffic, the slower it goes. To understand what is happening we must put this into mathematics.

A game theoretic model of travel times

An improved section of the road may take around 10 minutes to drive if you are alone on the road, compared to 30 minutes if the entire population of 200 drivers choose that road. A simple linear model for the time T to drive an improved section of the road in a certain direction is

$$T = 10 + \frac{x}{10},$$

where x denotes the number of drivers choosing to drive that section and in that direction. To avoid ambiguity we will typically use subscripts to denote the section and direction, such that x_{CD} refers to the number of drivers on the section from A-town to D-junction, and x_{DC} refers to the number of drivers on the same section in the opposite direction.

Under this model, travel times depend on how traffic is distributed on the possible roads. To predict how traffic will be distributed, we make the assumption that all travelers are trying to minimize their own travel time by choosing the quickest possible route. Which route is quickest depends on which routes other travelers choose. This dependence on what others do makes the choice of route a game theoretic problem.

Game theory is a branch of mathematics that studies the outcome of several interdependent individuals each trying to optimize their own behavior. In the terminology of game theory, every traveler is a *player*, a choice of route is a *strategy*, and the travel time is a *negative payoff* that each player wants to minimize. Assuming that all players successfully adapt their own strategies to minimize their own negative payoff, traffic will end up being distributed such that *each player's strategy is an optimal response to the other players' strategies*. Such a state is called a *Nash equilibrium* in game theory, after a 1950 paper by Nash [5]. In transportation science, the equilibrium state where no driver can improve her own travel time by switching route is usually called a *Wardrop user equilibrium*, after a 1952 paper by Wardrop [11]. The equivalence to the concept of the Nash equilibrium was realized later.

We shall now calculate how Nash equilibria changed as the A-town-B-town region's road system was developed.

Nash equilibrium in the original road system Before any improvements were made to the road system, players were not dependent on each others' choices. See Figure 1 (Left). To drive between A-town and B-town always took 120 ($= 60 + 60$) minutes, regardless of the traffic situation and whether you drove via C-junction or via D-junction. Both choices of routes were equally good, so both choices were optimal. Hence, any distribution of traffic on these two routes was a Nash equilibrium.

Nash equilibrium after improvement of road section AD Figure 1 (Right) shows the road system after improvement of the road between A-town and D-junction. An improved road section takes at most 30 minutes to drive, compared to 60 minutes for an old one. The improvement of road section AD therefore made the route via D-junction strictly superior to the route via C-junction. In other words, players had a unique optimal choice at this time: the route via D-junction. Consequently, the unique Nash equilibrium was that all 200 A-drivers and all 200 B-drivers would take the route via D-junction, yielding a travel time of $30 + 60 = 90$ minutes.

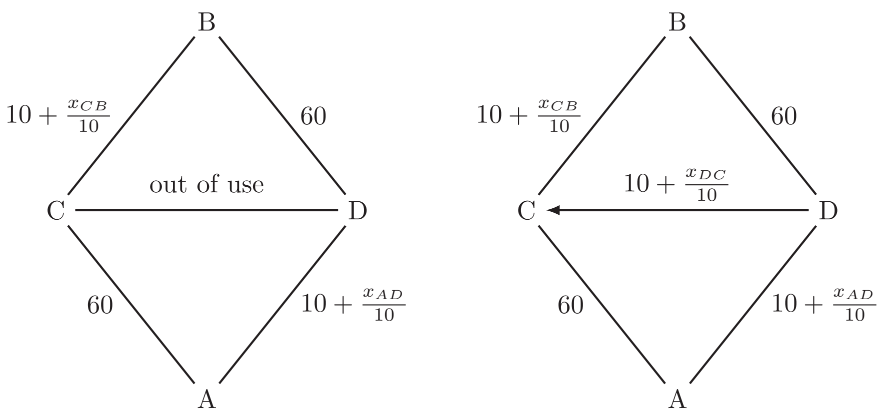


Figure 2 (Left) Travel times after improvement of road section CB. (Right) Travel times after improvement of road section DC.

Nash equilibrium after improvement of road section CB Figure 2 (Left) shows the road system after improvement of the road between C-junction and B-town. The competing routes via C-junction and via D-junction now have payoff functions of a similar form: $70 + \frac{x_{CB}}{10}$ and $70 + \frac{x_{AD}}{10}$, respectively. Whenever there were more traffic on one route, a driver on that route could shorten his or her travel time by choosing the other route instead. This means that drivers on the busier route were not making an optimal choice in response to the overall traffic situation. The only outcome in which every driver's choice was optimal was when both routes were equally busy.

Therefore, Nash equilibrium occurred when A-drivers divided themselves equally on the two routes, so that there were 100 A-drivers on each route (and the same for B-drivers). The travel time in this equilibrium is $70 + \frac{100}{10} = 80$ minutes.

Nash equilibrium after improvement of road section DC Figure 2 (Right) shows the road system after improvement of the road between D-junction and C-junction. As the new DC section is one-way, B-drivers could only use it together with the unimproved road sections—which would always be worse than their current routes. Thus, the traffic flow of B-drivers does not change. B-drivers' travel time therefore remains at 80 minutes.

By contrast, the traffic flow of A-drivers changes drastically. Drivers using the road section AC (60 minutes) would now always achieve a shorter (or at most equally long) drive by switching to the alternative road from A-town to C-junction via D-junction ($20 + \frac{x_{AD}+x_{DC}}{10}$, where $x_{AD}, x_{DC} \leq 200$). Similar reasoning applies to the road section DB. Thus, for all 200 drivers, the route that used all three improved road sections (AD, DC, and CB) is now an optimal choice. The travel time in this Nash equilibrium is $3 \cdot (10 + \frac{200}{10}) = 90$ minutes, which amounts to 10 minutes slower travel than before the road improvement.

Note that the last driver to switch is actually indifferent to switching and could just as well remain on the old route. Therefore, there is also a Nash equilibrium in which only 199 drivers use the crossroad and achieve a slightly shorter travel time of $3 \cdot 10 + \frac{200}{10} + \frac{199}{10} + \frac{199}{10} = 89.8$ minutes. The difference of 12 seconds to the other equilibrium is so marginal as to be practically irrelevant. For this reason a modeller may want to avoid having to deal with such phenomena by instead adopting a continuous traffic model, see the exercises at the end of this paper.

A social dilemma The above calculations give an explanation for the development of travel times that puzzled the council chairman. Individual drivers adjusted the routes they took such that they, as a group, used the entire road system less efficiently. Note that nothing forbids the A-drivers from going back to driving the old routes. If half the A-drivers would go via C-junction and the other half would go via D-junction, it would cut 10 minutes off everyone's travel time. The problem is that this efficient use of the road system is not a Nash equilibrium. Every individual A-driver would always be better off by deviating and choosing the ADCB route instead. To illustrate, consider the first driver to switch from the ADB route to the ADCB route. That driver would shorten her travel time from 80 minutes to $3 \cdot 10 + \frac{100}{10} + \frac{1}{10} + \frac{101}{10} = 50.2$ minutes. At the same time, that driver would slow down the other 100 drivers on the CB route by 0.1 minutes. By deviating, a driver gains time at the cost of others losing time. A similar calculation applies to a driver switching from the ACB route to the ADCB route.

Already Wardrop [11] pointed out that in a congested network, the user equilibrium will not necessarily minimize average travel time. The increase in average travel time due to individuals optimizing their own individual travel times is sometimes called the *price of anarchy*. Situations like these, where there is a conflict of interests between the individual and the group, are known in the game theory literature as *social dilemmas*. Other real-life examples range from the small scale (e.g., keeping a communal kitchen tidy) to the grand scale (e.g., curbing global carbon dioxide emissions) [10].

The Braess paradox The paradoxical phenomenon that making one road faster to travel can lead to slower travel overall (and vice versa) is called the *Braess paradox*. It was first demonstrated in 1968 by Dietrich Braess [2]. The Braess paradox is not just a theoretical curiosity. Indeed, according to mathematical analysis of random networks, the Braess paradox is very likely to occur [9]. It has been documented in many actual road systems across the world, such as when the traffic flow on Manhattan improved from a temporary closing of the 42nd street [4]. Analyses of real-world traffic networks in cities like San Francisco and Winnipeg have indicated that such *Braess links*, which decrease the overall performance of the network, may be quite common [1, 8]. Moreover, it is known that designing a network that is free from Braess links is a computationally hard problem [7].

The reason that the Braess paradox is so likely to occur in a general network is that it may arise whenever adding a new link causes some drivers to increase overall congestion. In the simple network in the example, it will typically arise whenever links CB and AD suffer from congestion (and thus are sensitive to traffic volumes), while road sections AC and DB do not. See exercise 3 at the end of the paper.

Note that the Braess paradox can occur in any kind of network with limited capacity. Examples include electrical networks [4] as well as computer networks, for which the Braess paradox is highly relevant due to the shortage of internet bandwidth and the strive to improve internet capacity [3].

In the remainder of this article we shall present a novel variation of the Braess paradox in road systems.

Incorporating meeting traffic

We are back at a council meeting, where the chairman has a harried look. He has received a strongly worded petition from the people living in B-town. They point out how unfair it is that the new road between C-junction and D-junction is open only for one-way traffic. As the road is paid for by the taxpayers, the petitioners demand it should be accessible to all morning rush drivers.

The budget does not allow construction of a second lane. "What can we do?", asks the chairman. Only one hand is raised. The chairman lets out a sigh: "Mrs. Jones." An excited Mrs. Jones suggests that they should keep the road a single lane and nonetheless open it for two-way traffic. "Ridiculous," says the chairman, "have you any idea how slow it is to drive on a single lane road with meeting traffic? If you think cars driving in the same direction slows traffic down, I would say meeting cars is ten times worse." Mrs. Jones answers the chairman with a nod and a smile.

The game of Chicken To incorporate the chairman's assumption that meeting cars would be ten times worse than having cars driving in the same direction, we assume that the payoff function, as shown in Figure 3 (Left), on the two-way road is

$$T_{DC} = 10 + \frac{x_{DC}}{10} + x_{CD} \text{ for cars driving from D-junction to C-junction, and}$$

$$T_{CD} = 10 + \frac{x_{CD}}{10} + x_{DC} \text{ for cars driving in the opposite direction.}$$

These payoff functions say that each car met adds another minute to the travel time.

Now consider a morning rush with drivers from A-town driving up the first improved road to D-junction, and drivers from B-town driving down the second improved road to C-junction. Both sides then face the choice of whether to use the new single-lane crossroad or the old roads. The worst outcome is if both sides were to choose the crossroad, as traffic would be excruciatingly slow. However, if all drivers coming from A-town yield by taking the old road instead, then drivers coming from B-town will happily take the crossroad, and vice versa.

For a game theorist this is reminiscent of a well-known game called Chicken (or Hawk-Dove), which models a conflict over a desirable resource where two agents face the following strategic choice: if your opponent yields then you should take the resource, but if your opponent does not yield then you are better off yielding than fighting over the resource. Because either agent can be the one that yields, the Chicken game has two pure Nash equilibria (as well as a "mixed" equilibrium in which both agents randomize whether to yield or not, see any book on game theory, e.g. [6]). In our road system, the new crossroad constitutes the desirable resource and we can view the collective of A-drivers and the collective of B-drivers as the opposing sides in the conflict over this resource.

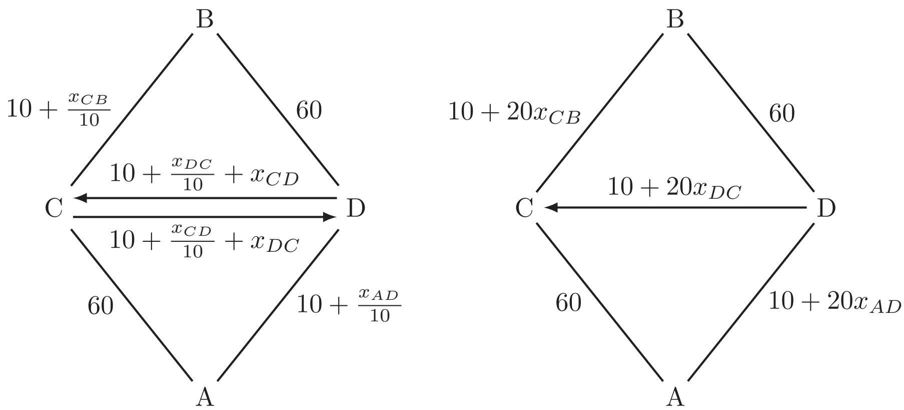


Figure 3 (Left) Travel times after opening road section DC for two-way traffic. (Right) Travel times after improvement of road section DC, after rescaling of traffic flow variables by a factor 200.

A new paradox

When the council meets again the chairman is glowing with joy. Since they opened the road between C-junction and D-junction for two-way traffic, two things have happened. When the B-drivers learned of the two-way solution, they agreed with their employers in A-town to start a little earlier in the morning in order to be first on the crossroad. And, paradoxically, the chairman's own morning drive in the opposite direction, from A-town to B-town, has quickened from 90 to 80 minutes. He smiles at Mrs. Jones and asks her to explain this new paradox: "How is it possible that my journey went quicker after we allowed meeting traffic?" Mrs. Jones starts by asking him if he actually drives the crossroad. "No, the B-drivers are obviously committed to using the crossroad, so now I just avoid it and drive one of the old roads instead."

Nash equilibrium after making road section DC two-way When the B-drivers commit to using the crossroad, A-drivers do best to yield, that is, to avoid the crossroad and use the old roads instead. B-drivers are then in the same situation as the A-drivers used to be, so it is a Nash equilibrium for the B-drivers to continue to use the crossroad. On the other hand, the A-drivers are in a situation equivalent to what they were in before the crossroad was improved. It is therefore a Nash equilibrium for A-drivers to adjust their routes such that they split into one half going via C-junction and the other half going via D-junction. The travel time for A-drivers is then back at 80 minutes, whereas the travel time for B-drivers increases to 90 minutes.

Note that every driver is still making an optimal response to the traffic situation. In particular, every B-driver's choice of route is optimal. What makes a B-driver's journey slow is the congestion created by the other B-drivers. Also note that a solution for the B-drivers would be to collectively decide to be the yielding side in the Chicken game, by renegotiating with their employers in A-town to start a little later in the morning instead. When A-drivers reach D-junction in the morning and observe the absence of meeting traffic on the crossroad, the individually optimal choice for them is then to switch back to using the crossroad. The situation would then be reversed, with A-drivers increasing their travel time to 90 minutes and B-drivers happily finding themselves in an equilibrium where their travel time has decreased back to 80 minutes.

The Chicken game arises when A-drivers or B-drivers collectively decide on a strategy with respect to use of the crossroad, either to go for it or to yield. If no such collective decision is made we can also obtain an equilibrium where the crossroad is used by a blend of a few A-drivers and a few B-drivers, see the exercises at the end.

The Chicken Braess paradox The new paradox is novel as far as we know. We call it the *Chicken Braess paradox*, as it incorporates both the game of Chicken and the Braess paradox. The Chicken Braess paradox demonstrates how yielding in a Chicken game can be a winning move in a broader context. It also shows how meeting traffic can make travel times shorter, which is a new variation on the Braess paradox.

Conclusion As we reviewed earlier, variations on the Braess paradox arise in real networks. It is important to be aware of this type of phenomenon when one seeks to improve networks, be they traffic systems, electrical grids, or computer networks such as the Internet. The takeaway for the aspiring network designer is to adopt the attitude of Mrs. Jones: be wary of the naïve intuition that improving or adding a link in a network will always improve overall network performance—and don't be afraid of using mathematics!

Exercises To further your insights into this problem, try the following exercises. As mentioned earlier, making the traffic model continuous may simplify the game theoretic analysis. In a continuous model we could represent the number of people traveling a certain section by a real number $x \in [0, 1]$, where a value of 1 represents 100% of the population of A-drivers (or the population of B-drivers). This means that traffic flow numbers are downscaled by the population size. Figure 3 (Right) shows how Figure 2 (Right) would be redrawn in this model. Use this continuous model for all the exercises.

(1) In the continuous model, show that the Nash equilibrium after improvement of road section DC is unique.

(2) After rescaling of traffic flows, the travel time on the crossroad when it is opened for two-way traffic is $10 + 20x_{CD} + 200x_{DC}$ in the CD-direction and $10 + 20x_{DC} + 200x_{CD}$ in the DC-direction. In the story we encountered equilibria where only A-drivers or only B-drivers use the crossroad, but there is also an equilibrium where the crossroad is used by a blend of a few A-drivers and a few B-drivers. To find this equilibrium, find the values of the traffic flow variables that satisfy that A-drivers have the same travel time along routes AC and ADC, and the same travel time along routes DB and DCB, and similarly for B-drivers. If you do it correctly, you will find a unique blended equilibrium in which all drivers have a total travel time of $81\frac{7}{23}$ minutes—which means that everyone loses compared to the situation before the crossroad was improved.

(3) The paradoxes do not depend on the exact numbers used in our example, nor on

travel times being linear functions of traffic. To generalize, let $T_{AD}(x)$ denote the travel time along road section AD when the traffic on this road section in this direction is x, and similarly for other road sections. Assume that the travel time functions are continuous and monotonically increasing (i.e., never decreasing) with traffic.

(a) Before the crossroad is improved we want both roads to be used in equilibrium. Explain why this holds if and only if $T_{AD}(1) + T_{DB}(1) > T_{AC}(0) + T_{CB}(0)$ and $T_{AD}(0) + T_{DB}(0) < T_{AC}(1) + T_{CB}(1)$, in which case there must exist an equilibrium traffic distribution given by $\hat{x} \in (0, 1)$ such that $T_{AD}(\hat{x}) + T_{DB}(\hat{x}) = T_{AC}(1 - \hat{x}) + T_{CB}(1 - \hat{x})$.

(b) In the story, improvement of the crossroad led to a new equilibrium in which all A-drivers used the crossroad. Explain why this happens if $T_{AD}(1) + T_{DC}(1) < T_{AC}(0)$ and $T_{DC}(1) + T_{CB}(1) < T_{DB}(0)$, and why we then have the Braess paradox if and only if $T_{AD}(1) + T_{DC}(1) + T_{CB}(1) > T_{AD}(\hat{x}) + T_{DB}(\hat{x})$. Finally, after opening the crossroad for two-way traffic, let $\overleftarrow{T}_{DC}(x)$ denote the time to travel the crossroad given that all B-drivers travel the crossroad in the opposite direction. Explain why the Chicken Braess paradox is obtained if $T_{AD}(\hat{x}) + \overleftarrow{T}_{DC}(0) > T_{AC}(1 - \hat{x})$ and $\overleftarrow{T}_{DC}(0) + T_{CB}(1 - \hat{x}) > T_{DB}(\hat{x})$.

(c) To illustrate the flexibility allowed by the conditions stated in exercise 3b, let's return to the special case described in Figure 3 (Right). Holding all other parameters constant, examine how we can vary the choices of travel time functions for the crossroad (i.e., $T_{DC}(x)$ and $\overleftarrow{T}_{DC}(x)$) and still obtain a new equilibrium in which all A-drivers use the crossroad and in which both the Braess paradox and the Chicken Braess paradox arise.

(d) Note that for the Braess paradox to arise it is not necessary that all A-drivers use the crossroad in the new equilibrium. Consider the generic case where the new equilibrium involves some A-drivers using the ACB route, some using the ADB route, and some using the ADCB route. In this case, explain why the Braess paradox will arise if the decrease in travel time from less traffic on road section AC is more than offset by the increase in travel time from more traffic on road section CB, and similarly for road sections DB and AD. For instance, this must hold when, as in Figure 2 (Right), travel times are sensitive to the amount of traffic on road sections CB and AD but not on road sections AC and DB.

REFERENCES

[1] Bagloee, S., Ceder, A., Tavana, M., Bozic, C. (2014). A heuristic methodology to tackle the Braess Paradox detecting problem tailored for real road networks. *Transportmetrica A*. 10(5): 437–456.

[2] Braess, D. (1968). ´Uber ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* 12(1): 258–268.

[3] Nagurney, A., Parkes, D., Daniele, P. (2007) The Internet, evolutionary variational inequalities, and the time-dependent Braess paradox. *Comput. Manag. Sci.* 4(4): 355–375.

[4] Nagurney, L., Nagurney, A. (2016). Physical proof of the occurrence of the Braess Paradox in electrical circuits. *EPL* 115(2): 28004.

[5] Nash, J. (1950). Equilibrium points in *n*-person games. *Proc. Natl. Acad. Sci. USA* 36(1): 48–49.

[6] Osborne, M., Rubinstein, A. (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.

[7] Roughgarden, T. (2001). Designing networks for selfish users is hard. In: *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science*. Los Alamitos, CA: IEEE Computer Society Press, pp. 472–481.

[8] Sun, L., Liu, L., Xu, Z., Jie, Y., Wei, D., Wang, P. (2015). Locating inefficient links in a large-scale transportation network, *Phys. A.* 419: 537–545.

[9] Valiant, G., Roughgarden, T. (2010). Braess's Paradox in large random graphs. *Random Struct. Algor.* 37(4): 495–515.

[10] Van Lange, P., Balliet, D., Parks, C., Van Vugt, M. (2014). *Social Dilemmas: Understanding Human Cooperation*. Oxford: Oxford University Press.

[11] Wardrop, J. (1952). Some theoretical aspects of road traffic research. *Proc. Inst. Civil Eng.* 1: 325–362.

Summary. The Braess Paradox is the counterintuitive fact that creation of a shortcut may make travel slower. As each driver seeks to minimize his/her travel time, the shortcut may become so popular that it causes congestion elsewhere in the road network, thereby increasing the travel time for everyone. We extend the paradox by considering a shortcut that is a single-lane but two-way street. The conflict about which drivers get to use the single-lane shortcut is an example of a game theoretic situation known as Chicken, which merges with the Braess Paradox into the novel Chicken Braess Paradox: meeting traffic may make travel quicker.

KIMMO ERIKSSON (MR Author ID: 292233) is a professor of mathematics at Mälardalen University in Västerås, Sweden. He took a Ph.D. in mathematics in 1993 and a second Ph.D. in social psychology 25 years later. One of his current research interests is why game theory is often a bad modelling paradigm; the present paper is an unusual example of the opposite.

JONAS ELIASSON (MR Author ID: 1296267) is a former professor of transport systems analysis at the Royal Institute of Technology. A certain frustration with the inability of decision-makers to grasp the finer details of transportation research led to him to accept the position as director of the Stockholm Transport Administration, which has led to a certain frustration with the inability of transport researchers to grasp the finer details of real-world decision-making.

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

Two Convergence Theorems and an Extension of the Ratio Test for a Series

ZILI WU
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, 215123 China
ziliwu@email.com

Introduced in the calculus sequence, the ratio test plays an important role in determining the convergence of a series $\sum_{n=1}^{\infty} a_n$ of nonnegative terms. Recall that if

$$\lim_{n\to+\infty} \frac{a_{n+1}}{a_n} = r,$$

then the series converges if $r < 1$, diverges if $r > 1$, and is inconclusive if $r = 1$. One such example when the ratio test is not applicable is for a p-series for which $a_n = \frac{1}{n^p}$. In this paper, we consider alternate convergence tests, which will be able to determine the convergence for p-series.

We recall that the terms of an absolutely convergent series are allowed to be rearranged without changing its convergence and value. So a sequence $\{a_n\}$ of nonnegative terms can be assumed to be decreasing when the convergence of $\sum_{n=1}^{\infty} a_n$ is considered. This is one of the main cases where several convergence tests are applied. For such a series, the Cauchy condensation test states that $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{n=1}^{\infty} a_n b_n$ does, where $b_n = 2^k$ for $n = 2^k$ and $b_n = 0$ elsewhere. This result has been extended to the case where the sequence $\{b_n\}$ of nonnegative terms is *effective* for monotone series, that is, for every decreasing sequence $\{a_n\}$ of nonnegative terms the series $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} a_n b_n$ have the same convergence. With some complicated concepts and techniques in functional analysis, it has been shown that $\{b_n\}$ is effective for monotone series if and only if it satisfies

$$0 < \liminf_{n\to+\infty} \frac{b_1 + \cdots + b_n}{n} \leq \limsup_{n\to+\infty} \frac{b_1 + \cdots + b_n}{n} < \infty$$

(see [1]). From this we see that $\sum_{n=1}^{\infty} b_n$ must be divergent. In practice, to discuss the convergence of $\sum_{n=1}^{\infty} a_n$, it is not necessary to use all effective sequences, only some. Note that the above equivalent statement for an effective sequence contains the partial sums of $\sum_{n=1}^{\infty} b_n$ with the limit inferior and the limit superior of a relevant sequence, which are not always easy to be found. Moreover, to discuss the convergence of $\sum_{n=1}^{\infty} a_n b_n$, we may still need to apply other convergence tests!

Even though there exist several convenient convergence tests for series, we still have lots of series whose convergence cannot be easily discussed by these tests. As we stated above, the convergence of a p-series cannot be discussed only by the ratio test. Therefore it makes sense to investigate some simpler convergence tests. In this paper we mainly present several cases for which equivalent statements for the convergence of $\sum_{n=1}^{\infty} a_n$ can conveniently be established by simple, natural, and self-contained proofs. For $1 < q \in \mathbb{N}$, we prove that $\sum_{k=1}^{\infty} q^k a_{q^k}$ and $\sum_{n=1}^{\infty} a_n$ have the same convergence. With the convergence of $\sum_{k=1}^{\infty} q^k a_{q^k}$ and the ratio test, a more convenient ratio test is also achieved. One example further shows that the new extended ratio test is applicable to more series of nonnegative terms than the ratio test.

Math. Mag. **92** (2019) 222–227. doi:10.1080/0025570X.2019.1560841 © Mathematical Association of America
MSC: Primary 40A05

Two convergence theorems

Based on the monotonicity of $\{a_n\}$, our first theorem states that the convergence of $\sum_{n=1}^{\infty} a_n$ is equivalent to the convergence of a condensed series of the form $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_k}$.

Theorem 1. *Let $\{a_n\}$ satisfy $a_n \geq a_{n+1} \geq 0$ for all $n \in \mathbb{N}$. Then the following are equivalent:*

(i) *$\sum_{n=1}^{\infty} a_n$ converges.*

(ii) *For each subsequence $\{n_k\}$ of $\{n\}$, $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_{k+1}}$ converges.*

(iii) *For any $M > 0$, if a subsequence $\{n_k\}$ of $\{n\}$ satisfies*

$$(n_{k+1}-n_k) \leq M(n_k - n_{k-1}) \quad \textit{for all } 1 < k \in \mathbb{N},$$

then $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_k}$ converges.

(iv) *For any $M \geq 1$ there exists a subsequence $\{n_k\}$ of $\{n\}$ such that*

$$1 < n_{k+1}-n_k < (M+1)(n_k - n_{k-1}) \quad \textit{for all } 1 < k \in \mathbb{N}$$

and $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_k}$ converges.

(v) *For any $M \geq 1$, if a subsequence $\{n_k\}$ of $\{n\}$ satisfies $n_{k+1}-n_k \leq M$ for all $k \in \mathbb{N}$, then $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_k}$ converges.*

(vi) *If $n_1 \in \mathbb{N}$ and $d \in \mathbb{N}$, then $\sum_{k=0}^{\infty} a_{n_1+kd}$ converges.*

Proof. We show $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i)$ and $(iii) \Rightarrow (v) \Rightarrow (vi) \Rightarrow (i)$.

$(i) \Rightarrow (ii)$: Let $\sum_{n=1}^{\infty} a_n$ converge. Then $\{s_n\}$ is bounded, where

$$s_n = a_1 + a_2 + \cdots + a_n \quad \text{for } n \in \mathbb{N}.$$

For any subsequence $\{n_k\}$ of $\{n\}$, denote

$$t_m = n_1 a_{n_1} + (n_2 - n_1)a_{n_2} + \cdots + (n_{m+1}-n_m)a_{n_{m+1}} \quad \text{for } m \in \mathbb{N}.$$

For any $n_{m+1} \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that $n_{m+1} < n$. Since $\{a_n\}$ is decreasing,

$$\begin{aligned} s_n &\geq (a_1 + \cdots + a_{n_1}) + (a_{n_1+1} + \cdots + a_{n_2}) + \cdots + (a_{n_m+1} + \cdots + a_{n_{m+1}}) \\ &\geq n_1 a_{n_1} + (n_2 - n_1)a_{n_2} + \cdots + (n_{m+1}-n_m)a_{n_{m+1}} = t_m \geq 0. \end{aligned}$$

Thus $\{t_m\}$ is bounded since $\{s_n\}$ is. It follows that $\{t_m\}$ converges. This implies that $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_{k+1}}$ converges.

$(ii) \Rightarrow (iii)$: Suppose that (ii) is true. Then, for any $M > 0$, if a subsequence $\{n_k\}$ of $\{n\}$ satisfies, for all $1 < k \in N$,

$$(n_{k+1}-n_k) \leq M(n_k - n_{k-1}), \quad \text{then} \quad (n_{k+1}-n_k)a_{n_k} \leq M(n_k - n_{k-1})a_{n_k}$$

for all $1 < k \in \mathbb{N}$. Since $\sum_{k=2}^{\infty}(n_k - n_{k-1})a_{n_k}$ converges, so does $\sum_{k=2}^{\infty} M(n_k - n_{k-1})a_{n_k}$. Thus, by the comparison test, $\sum_{k=1}^{\infty}(n_{k+1}-n_k)a_{n_k}$ converges.

$(iii) \Rightarrow (iv)$: Let $M \geq 1$. Then, for any given $n_1, n_2 \in \mathbb{N}$ with $1 < n_2 - n_1$, there exists $n_3 \in \mathbb{N}$ such that $n_3 \leq M(n_2 - n_1) + n_2 + 1 < n_3 + 1$, from which it follows that $1 < n_3 - n_2 < (M+1)(n_2 - n_1)$.

Suppose that $n_k, n_{k+1} \in \mathbb{N}$ satisfy $1 < n_{k+1} - n_k < (M+1)(n_k - n_{k-1})$. Then there exists $n_{k+2} \in \mathbb{N}$ such that $n_{k+2} \leq M(n_{k+1}-n_k) + n_{k+1} + 1 < n_{k+2} + 1$. Based on this we have $1 < n_{k+2} - n_{k+1} < (M+1)(n_{k+1}-n_k)$. Therefore, by induction, such a sequence $\{n_k\}$ satisfies $1 < n_{k+1} - n_k < (M+1)(n_k - n_{k-1})$ for all $1 < k \in \mathbb{N}$. It follows from (iii) that $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges.

$(iv) \Rightarrow (i)$: Suppose that (iv) holds. Then there exists a subsequence $\{n_k\}$ of $\{n\}$ such that $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges. This is equivalent to saying that the series $\{\sum_{k=1}^{m}(n_{k+1} - n_k)a_{n_k}\}$ is bounded. Since for $n_1 < m \in \mathbb{N}$ we have

$$0 \le \sum_{n=1}^{m} a_n \le \sum_{n=1}^{n_1} a_n + \sum_{k=1}^{m}(n_{k+1} - n_k)a_{n_k},$$

$\{\sum_{n=1}^{m} a_n\}$ is bounded. Thus (i) follows.

$(iii) \Rightarrow (v)$: Let (iii) be true. Then, for $M \ge 1$, if a subsequence $\{n_k\}$ of $\{n\}$ satisfies $n_{k+1} - n_k \le M$ for all $k \in N$, then

$$(n_{k+1} - n_k) \le M(n_k - n_{k-1}) \qquad \text{for all } 1 < k \in N.$$

Thus, by (iii), $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges.

The implication $(v) \Rightarrow (vi)$ is trivial by taking $n_k = n_1 + (k-1)d$ and $M = d$ in (v) while $(vi) \Rightarrow (i)$ follows easily by taking $n_1 = 1$ and $d = 1$ in (vi). Therefore the proof is complete. ∎

Based on Theorem 1, if there exists a subsequence $\{n_k\}$ of $\{n\}$ such that the series $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_{k+1}}$ diverges, then from the implication $(i) \Rightarrow (ii)$ the series $\sum_{n=1}^{\infty} a_n$ must be divergent. For example, for the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$, since $a_n = \frac{1}{n}$, if $n_k = k!$, then

$$\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_{k+1}} = \sum_{k=1}^{\infty}[(k+1)! - k!]\frac{1}{(k+1)!} = \sum_{k=1}^{\infty}\frac{k}{k+1},$$

which is divergent. Therefore $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges.

We can also prove this conclusion by contradiction. Indeed, suppose that $\sum_{n=1}^{\infty} \frac{1}{n}$ converges to some $a \in \mathbb{R}$. Then, by (vi) of Theorem 1, both $\sum_{k=0}^{\infty} \frac{1}{2k+1}$ and $\sum_{k=1}^{\infty} \frac{1}{2k}$ converge, which leads to the following contradiction:

$$a = \sum_{n=1}^{\infty}\frac{1}{n} = \sum_{k=0}^{\infty}\frac{1}{2k+1} + \sum_{k=1}^{\infty}\frac{1}{2k} > 2\sum_{k=1}^{\infty}\frac{1}{2k} = \sum_{k=1}^{\infty}\frac{1}{k} = a.$$

By the equivalence $(i) \Leftrightarrow (ii)$ in Theorem 1 and a necessary condition for a convergent series, we obtain the following divergence test immediately.

Corollary 1. *For a decreasing sequence $\{a_n\}$ of nonnegative terms, if there exists a subsequence $\{n_k\}$ of $\{n\}$ such that $(n_{k+1} - n_k)a_{n_{k+1}} \not\to 0$ as $k \to +\infty$, then $\sum_{n=1}^{\infty} a_n$ diverges.*

From (iv) and (v) of Theorem 1, it suffices to consider the convergence of $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$—instead of $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_{k+1}}$ stated in (ii)—for a subsequence $\{n_k\}$ satisfying

$$1 < n_{k+1} - n_k < (M+1)(n_k - n_{k-1}) \qquad \text{for some } M \ge 1 \text{ and all } 1 < k \in \mathbb{N}$$

to determine the convergence of $\sum_{n=1}^{\infty} a_n$. In addition, if $\{n_{k+1} - n_k\}$ is bounded, then the factor $(n_{k+1} - n_k)$ in the general term of $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ can be removed. However, if $\{n_{k+1} - n_k\}$ is unbounded, the factor is still needed since the convergence of $\sum_{k=1}^{\infty} a_{n_k}$ may be different from that of $\sum_{n=1}^{\infty} a_n$. For example, for $\sum_{n=1}^{\infty} \frac{1}{n}$, if we take $n_k = k^2$, then

$$\sum_{k=1}^{\infty} \frac{1}{n_k} = \sum_{k=1}^{\infty} \frac{1}{k^2}$$

is convergent but $\sum_{n=1}^{\infty} \frac{1}{n}$ is not.

From Theorem 1, the convergence of $\sum_{n=1}^{\infty} a_n$ implies the existence of a subsequence $\{n_k\}$ of $\{n\}$ such that $n_k \neq k$ for all $k \in \mathbb{N}$ and $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges. Conversely, as the proof for $(iv) \Rightarrow (i)$ of Theorem 1 shows, the series $\sum_{n=1}^{\infty} a_n$ must converge if the series $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges (no matter whether $\{n_{k+1} - n_k\}$ is bounded or not). Indeed, the convergence of $\sum_{n=1}^{\infty} a_n$ is completely characterized by the convergence of the series $\sum_{k=1}^{\infty} q^k a_{q^k}$ for $1 < q \in \mathbb{N}$, as the following theorem states.

Theorem 2. *Let $\{a_n\}$ satisfy $a_n \geq a_{n+1} \geq 0$ for all $n \in \mathbb{N}$. Then for any $1 < q \in \mathbb{N}$ the following are equivalent:*

- (i) *$\sum_{n=1}^{\infty} a_n$ converges.*
- (ii) *$\sum_{k=1}^{\infty} q^k a_{q^k}$ converges.*
- (iii) *There exists a subsequence $\{n_k\}$ of $\{n\}$ such that $\{n_{k+1} - n_k\}$ is increasing and both $\sum_{k=1}^{\infty} \frac{1}{n_{k+1}-n_k}$ and $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converge.*
- (iv) *There exists a subsequence $\{n_k\}$ of $\{n\}$ such that $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges.*

Proof. Let $1 < q \in \mathbb{N}$. We prove $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i)$.

$(i) \Rightarrow (ii)$: If $\sum_{n=1}^{\infty} a_n$ converges, then, by the implication $(i) \Rightarrow (iii)$ of Theorem 1, for $n_k = q^k$, the series

$$\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k} = (q-1)\sum_{k=1}^{\infty} q^k a_{q^k}$$

converges since $(n_{k+1} - n_k) = q(q^k - q^{k-1}) = q(n_k - n_{k-1})$ for all $1 < k \in \mathbb{N}$. Thus (ii) follows.

$(ii) \Rightarrow (iii)$: For $n_k = q^k$, the sequence $\{n_{k+1} - n_k\}$ is increasing and the series $\sum_{k=1}^{\infty} \frac{1}{n_{k+1}-n_k} = \sum_{k=1}^{\infty}(q-1)^{-1}q^{-k}$ is a convergent geometric series. In addition, by (ii),

$$\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k} = (q-1)\sum_{k=1}^{\infty} q^k a_{q^k}$$

converges. So (iii) is valid.

$(iii) \Rightarrow (iv)$ is obvious, so it remains to show $(iv) \Rightarrow (i)$.

$(iv) \Rightarrow (i)$: Let $\{n_k\}$ be a subsequence of $\{n\}$ such that $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ converges. Then for any $m \in \mathbb{N}$ with $n_1 < m$ we have

$$0 \leq s_m := a_1 + a_2 + \cdots + a_m \leq n_1 a_1 + \sum_{k=1}^{m}(n_{k+1} - n_k)a_{n_k}.$$

This implies that $\{s_n\}$ is bounded due to the convergence of $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$. So $\sum_{n=1}^{\infty} a_n$ is convergent. ∎

For a decreasing sequence $\{a_n\}$ of nonnegative terms, by Theorem 2, $\sum_{n=1}^{\infty} a_n$ diverges if and only if $\sum_{n=0}^{\infty} q^n a_{q^n}$ diverges for some $1 < q \in \mathbb{N}$. In particular, based on a necessary condition for a convergent series, we have the following conclusion.

Corollary 2. *For a decreasing sequence $\{a_n\}$ of nonnegative terms, if $q^k a_{q^k} \not\to 0$ as $k \to +\infty$, then $\sum_{n=1}^{\infty} a_n$ diverges.*

As an application of Corollary 2, it is immediate to see that the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent since for $a_n = \frac{1}{n}$ we have $q^k a_{q^k} = 1 \not\to 0$ as $k \to +\infty$.

Remark. (a) For a subsequence $\{n_k\}$ in Theorems 1 and 2, if we take $b_n = n_{k+1} - n_k$ for $n = n_k$ and $b_n = 0$ elsewhere, then $\{b_n\}$ is an effective sequence for monotone series, so these two theorems can also follow directly from [**1**]. However, as we have seen above, our proofs do not need functional analysis. In addition, these two theorems are easier to apply.

(b) Theorem 2 is motivated by the Cauchy condensation test which can in turn be obtained from Theorem 2 by taking $q = 2$.

(c) For a sequence $\{a_n\}$ with $a_n \ge a_{n+1} \ge 0$ for all $n \in \mathbb{N}$ in Theorem 2, the convergence of $\sum_{n=1}^{\infty} a_n$ implies that for the subsequence $\{n_k\}$ of $\{n\}$ in (iii), we have $\frac{a_{n_k}}{\frac{1}{n_{k+1}-n_k}} \to 0$ as $k \to +\infty$. In particular, for $n_k = q^k$, this limit reduces to $\frac{(q-1)a_{q^k}}{q^{-k}} \to 0$ as $k \to +\infty$, that is, $a_{q^k} \to 0$ much faster than $q^{-k} \to 0$ as $k \to +\infty$. The larger the number q is, the faster the term a_{q^k} goes to 0 as $k \to +\infty$.

An extension to the ratio test

For a decreasing sequence $\{a_n\}$ of nonnegative terms, consider the series $\sum_{k=1}^{\infty} q^k a_{q^k}$ for $1 < q \in \mathbb{N}$. By the ratio test, if

$$\lim_{k\to+\infty} \frac{q^{k+1} a_{q^{k+1}}}{q^k a_{q^k}} = \lim_{k\to+\infty} \frac{q a_{q^{k+1}}}{a_{q^k}} < 1, \text{ that is, } \lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} < \frac{1}{q},$$

then $\sum_{k=1}^{\infty} q^k a_{q^k}$ converges. Similarly, if $\lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} > \frac{1}{q}$, then $\sum_{k=1}^{\infty} q^k a_{q^k}$ diverges. Thus, based on Theorem 2, we obtain the following test.

Theorem 3. *Let $1 < q \in \mathbb{N}$ and $a_n \ge a_{n+1} \ge 0$ for all $n \in \mathbb{N}$.*

(i) If $\lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} < \frac{1}{q}$, then $\sum_{n=1}^{\infty} a_n$ converges.

(ii) If $\lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} > \frac{1}{q}$, then $\sum_{n=1}^{\infty} a_n$ diverges.

Theorem 3 is an extension to the ratio test. Indeed, it can be applied to give a new proof of the ratio test via the following result.

Proposition. *Let $a_n > 0$ for all $n \in \mathbb{N}$ and $\rho := \lim\limits_{n\to+\infty} \frac{a_{n+1}}{a_n}$ exist. If $\rho < 1$, then for $1 < q \in \mathbb{N}$, the following limit holds:*

$$\lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} = 0.$$

Proof. If $\rho < 1$, then for any $0 < \epsilon < 1 - \rho$ there exists $n_0 \in \mathbb{N}$ such that

$$\frac{a_{n+1}}{a_n} < \rho + \epsilon < \frac{n_0}{1 + n_0} < 1 \quad \text{for all } n \ge n_0.$$

From this we have $0 \leq a_{n+1} < a_n$ for all $n \geq n_0$ and, for $1 < q \in \mathbb{N}$ and $k \geq n_0$,

$$0 \leq \frac{a_{q^{k+1}}}{a_{q^k}} = \frac{a_{q^k+1}}{a_{q^k}} \cdot \frac{a_{q^k+2}}{a_{q^k+1}} \cdot \cdots \cdot \frac{a_{q^{k+1}}}{a_{q^{k+1}-1}} < \left[\frac{n_0}{1+n_0}\right]^{(q-1)q^k}.$$

This implies that $\lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} = 0.$ ■

Since the ratio test can be obtained by Theorem 3, convergence of $\sum_{n=1}^{\infty} a_n$ determined by the ratio test can also be determined by Theorem 3. However, its converse fails. There exist cases where the ratio test does not work but Theorem 3 does.

Example. To determine the convergence of the p-series $\sum_{n=1}^{\infty} \frac{1}{n^p} \quad (p > 0)$, let $a_n = \frac{1}{n^p}$ for $n \in \mathbb{N}$ and $1 < q \in \mathbb{N}$. Then

$$\lim_{k\to+\infty} \frac{a_{q^{k+1}}}{a_{q^k}} = \lim_{k\to+\infty} \left(\frac{q^k}{q^{k+1}}\right)^p = \frac{1}{q^p}.$$

By Theorem 3, the p-series is convergent for $p > 1$ and divergent for $0 < p < 1$.

For the case $p = 1$, the p-series is $\sum_{n=1}^{\infty} \frac{1}{n}$, which diverges as we know.

However, since

$$\rho = \lim_{n\to+\infty} \frac{a_{n+1}}{a_n} = \lim_{n\to+\infty} \left(\frac{n}{n+1}\right)^p = 1,$$

the ratio test is not applicable to the p-series.

Acknowledgments The author thanks the anonymous referees for their valuable comments, suggestions, and support! This paper is dedicated to Mr. Zhanghong Wu and Jingu High School, the 9th High School in Anxi, Fujian, China.

REFERENCES

[1] Alexiewicz, A. (1957). On Cauchy's condensation theorem. *Studia Math.* 16: 80–85.

Summary. Given a decreasing sequence $\{a_n\}$ of nonnegative terms, we prove that there exist subsequences $\{n_k\}$ of $\{n\}$ such that $\sum_{k=1}^{\infty}(n_{k+1} - n_k)a_{n_k}$ and $\sum_{n=1}^{\infty} a_n$ have the same convergence. In particular, $\sum_{n=1}^{\infty} a_n$ converges if and only if $\sum_{k=1}^{\infty} q^k a_{q^k}$ converges for any natural number $q > 1$. Using this result, the ratio test is also extended. One example further shows that the new extended ratio test is applicable to discussing the convergence of a p-series for which the ratio test fails.

ZILI WU (MR Author ID: 658450) received his Bachelor of Science from Xiamen University and Ph.D. from University of Victoria. Currently, he is an associate professor at Xi'an Jiaotong-Liverpool University. His research interests include error bounds for an inequality system, fixed point theory, variational principles, and variational inequalities.

ACROSS

1. (36, 42) = 6 and (24, 35) = 1, e.g.
5. * Mohamed of Harvey Mudd College who will give an invited address for students about graduate school
9. Praise
13. Garden tools
14. Scent in the air
15. Memo heading meaning "with respect to"
16. Bacardi and Captain Morgan
17. * Ami of Pomona College who will give an invited address about uncertainty
19. Hoppy beer: Abbr.
20. Like a mesh with small elements
21. Inverts, like a capsizing boat: ___ over
22. Measurement of the difference between a car's direction of motion and where its front wheels are pointing
25. ___ thick, as with praise
27. Basic belief (but not an axiom)
28. Annual accolade from MTV
30. Beatles hit: "___ Loves You"
31. Kind of lily that is Utah's state flower
32. Science fiction author Egan who recently made progress on an open problem about superpermutations
34. 54 for Xenon, e.g.: Abbr.
35. * Rochelle of Univ. of Illinois who will give an invited address about rehumanizing mathematics
38. Third-person plural pronoun
41. Dog in a classic Disney film
42. "Livin' La ___ Loca"
45. ___unit, number consisting of all 1's
46. British mathematician and popular-science writer Stewart
47. Like Xenon gas
49. Butts, to Brits
51. Fact in topology that is essential for proving that a finite product of compact spaces is compact
55. 10^3 in France
57. Actions at an auction
58. Big mo. for the IRS
59. * Alice of UC Irvine who will give an invited address about cryptography
61. Like a desert
62. Midwestern Native American tribe
63. TV network that airs Storage Wars
64. Prefix repeated in ___pus, ___genarian, ___mom
65. Units, ___, hundreds, . . .
66. Lays down the lawn
67. Defeat

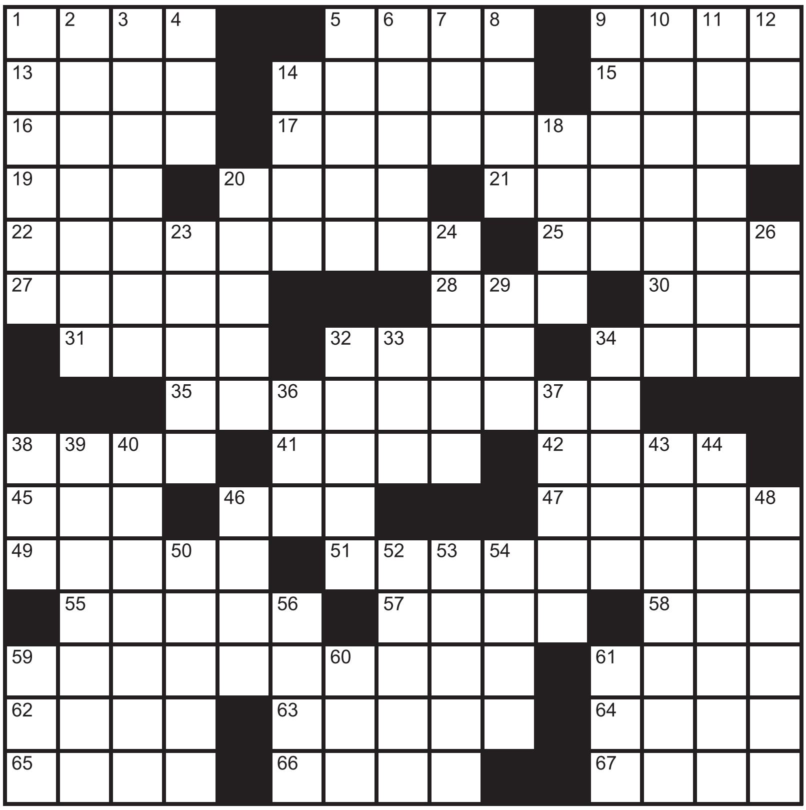
DOWN

1. * Robert of U. Penn. who will give an invited address about multivariable calculus
2. Variant of handshake problem involves married ___
3. * Erik of MIT who will give an invited address about Martin Gardner
4. Snake's sound
5. Primate at the zoo (for short)
6. Mathematical description of a real world system
7. Xenon measures 131.29 of these
8. Dimension of the column space of a matrix
9. "___ Rolling Stone"
10. Mathematician who studies measure theory, for instance
11. Soviet mathematician whose namesake result says a topological space is normal iff any two disjoint closed sets can be separated by a continuous function
12. Gov't agency combating smuggling
14. Org. that administers IP addresses
18. Actress Ward
20. French mathematician, namesake for the complement of a Julia set
23. American figure skater turned commentator Fleming
24. $\forall$ = for ___
26. "The Matrix" hero
29. Michael Scott's role at Dunder Mifflin Scranton: Abbr.
32. One of many that Newton stood on the shoulder of
33. * Baseball player in the host city of this year's MathFest
34. Organic compound with a ring of nitrogen atoms
36. This clue's answer is self-referential and could be the clue itself: Abbr.
37. "Lesser of two ___"
38. "___-la-la"
39. French mathematician who first proved that *e* is transcendental
40. * Math honor society with conference at MathFest: Pi Mu ___
43. * Laura of Northwestern Univ. who will give a series of invited addresses about elliptic curves
44. Axillae
46. Wight or Man
48. * Eva of Cornell Univ. who will give an invited address about games and Nash equilibria
50. Santa's helpers
52. Common shape of piping, named for the letter it resembles
53. Blue jays, parakeets, owls, etc.
54. It joins two vertices in a graph
56. Historic periods
59. Offensive slang for alcoholic
60. Vietnamese Fields Medalist Ngo ___ Chau
61. "You've got mail" co.

Math. Mag. **92** (2019) 228–229. doi:10.1080/0025570X.2019.1611155 © Mathematical Association of America

MathFest 2019

BRENDAN SULLIVAN
Emmanuel College
Boston, MA
sullivanb@emmanuel.edu



Clues start at left, on page 228. The Solution is on page 221.

Extra copies of the puzzle can be downloaded for free from the Taylor & Francis website https://www.tandfonline.com/umma.

Crossword Puzzle Creators

If you are interested in submitting a mathematically themed crossword puzzle for possible inclusion in MATHEMATICS MAGAZINE, please contact the editor at mathmag@maa.org.

PROBLEMS

Proposals

To be considered for publication, solutions should be received by November 1, 2019.

2071. *Proposed by Ioan Băetu, Botoşani, Romania.*

Let $n > 1$ be an integer, and let $\mathbb{Z}_n$ be the ring of integers modulo n. For fixed $k \in \mathbb{Z}_n - \{0\}$, define a binary operation "$\circ$" on $\mathbb{Z}_n$ by $x \circ y = (x-k)(y-k)+k$ for all $x, y \in \mathbb{Z}_n$. Let U be the group of units of $\mathbb{Z}_n$ (under multiplication), and let U_k° be the set of elements of $\mathbb{Z}_n$ that are invertible under the operation $\circ$. Characterize those n with the property that $U \neq U_k^\circ$ for all $k \in \mathbb{Z}_n - \{0\}$.

2072. *Proposed by Julien Sorel, Piatra Neamt, PNI, Romania.*

(a) Show that the initial value problem

$$\begin{cases} y' = \sqrt{1-y^2}, \\ y(0) = 1 \end{cases}$$

has infinitely many solutions defined on $\mathbb{R}$.

(b) By contrast, show that the initial value problem

$$\begin{cases} y' = \sqrt{x^2-y^2}, \\ y(1) = 1 \end{cases}$$

has no solutions defined on an open interval containing $x = 1$.

Math. Mag. **92** (2019) 230–238. doi:10.1080/0025570X.2019.1599245.

We invite readers to submit original problems appealing to students and teachers of advanced undergraduate mathematics. Proposals must always be accompanied by a solution and any relevant bibliographical information that will assist the editors and referees. A problem submitted as a Quickie should have an unexpected, succinct solution. Submitted problems should not be under consideration for publication elsewhere.

Proposals and solutions should be written in a style appropriate for this MAGAZINE.

Authors of proposals and solutions should send their contributions using the Magazine's submissions system hosted at http://mathematicsmagazine.submittable.com*. More detailed instructions are available there. We encourage submissions in PDF format, ideally accompanied by LaTeX source. General inquiries to the editors should be sent to* mathmagproblems@maa.org

2073. *Proposed by Enrique Treviño, Lake Forest College, Lake Forest, IL.*

A *factorial expansion* is any formal expression of the form

$$\overline{a_k a_{k-1} \dots a_2 a_1},$$

where $a_1, a_2, \dots, a_k$ are k integers ($k \geq 1$) such that $0 \leq a_i \leq i$ for $i = 1, 2, \dots, k$. The *value* of such a factorial expansion is

$$a_k \cdot k! + a_{k-1} \cdot (k-1)! + \cdots + a_2 \cdot 2! + a_1 \cdot 1!.$$

If the integers $a_1, \dots, a_k$ are expressed in base 10 and their digits simply written together without separation, the value of the factorial expansion so written is often ambiguous. For instance, the expansion $\overline{10000000000}$ may be interpreted as having coefficients 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 and value $1 \times 11! + 0 \times (10! + 9! + \cdots + 1!) = 11!$, or having coefficients 10, 0, 0, 0, 0, 0, 0, 0, 0, 0 and value $10 \times 10! + 0 \times (9! + 8! + 7! + \cdots + 1!) = 10 \times 10!$. Such factorial expansions are called *ambiguous.* On the other hand, some factorial expansions are unambiguous: for example, the expansion $\overline{311}$ must have the value $3 \times 3! + 1 \times 2! + 1 \times 1! = 21$. Prove that there are only finitely many unambiguous factorial expansions, and find the one whose value is largest.

2074. *Proposed by Bao Do (student), Columbus State University, Columbus, GA.*

Evaluate

$$\lim_{n\to\infty} \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} \binom{n}{k} H_k,$$

where $H_k = \sum_{j=1}^{k} \frac{1}{j}$ is the kth harmonic sum.

2075. *Proposed by Michael Goldenberg, The Ingenuity Project, Baltimore Polytechnic Institute, Baltimore, MD and Mark Kaplan, Towson University, Towson, MD.*

Consider the sequence $\{C_n\}$ defined recursively by $C_0 = 3$, $C_1 = 1$, $C_2 = 3$, and

$$C_n = C_{n-1} + C_{n-2} + C_{n-3} \qquad \text{for } n \geq 3.$$

Let $O = (0, 0, 0)$ be the origin of $\mathbb{R}^3$ and, for integer $n \geq 0$, let P_n be the point (C_n, C_{n+1}, C_{n+2}).

(a) Find the volume of the pyramid $O P_n P_{n+1} P_{n+2}$ in closed form.

(b) Show that the sequence $\{P_n\}$ asymptotically approaches a fixed line $\mathcal{L}$ through the origin of $\mathbb{R}^3$, and characterize this line.

Quickies

1091. *Proposed by H. A. ShahAli, Tehran, Iran.*

Show that no more than two straight cuts are needed to split any triangle into three or fewer pieces that may be rearranged (without overlap or gap) to make a right triangle.

1092. *Proposed by Ovidiu Furdui and Alina Sîntămărian, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.*

Given a positive integer n, evaluate

$$\int_0^1 \int_0^1 \arctan\left(\frac{x^n}{y^n}\right) dx\, dy.$$

Solutions

A characterization of S_m as subgroup of S_n for $m < n$ **June 2018**

2046. *Proposed by Ioan Băetu, Botoşani, Romania.*

For integers m, n such that $1 \le m < n$, let S_n be the group of all permutations of $\{1, 2, \dots, n\}$, let F be the set of permutations $\sigma \in S_n$ such that $\sigma(m) < \sigma(m+1) < \cdots < \sigma(n)$, and let T be the set of transpositions in F. Prove that there exists a unique subgroup G of S_n such that $T \subset G \subset F$.

Solution by Joseph DiMuro, Biola University, La Mirada, CA.
We show that the only subgroup G of S_n satisfying the given hypotheses is $G = S_m$ regarded as the set of permutations in S_n fixing each of $m+1, \dots, n$.

First, any transposition $\sigma \in S_m$ satisfies $\sigma(m) \le m$ and $\sigma(i) = i$ for all $i > m$, hence $\sigma(m) \le m < m+1 = \sigma(m+1) < \cdots < n = \sigma(n)$, so $\sigma \in T$. Therefore, T contains all transpositions in S_m; thus, G includes the group generated by those transpositions, which is S_m itself.

Conversely, we show that G contains no other permutations. Let $\tau \in S_n - S_m$ be arbitrary. Since $\tau \notin S_m$, τ is not the identity permutation, hence there exists $i \in \{1, \dots, n\}$ such that $\tau(i) \ne i$. Since $\tau \notin S_m$, the largest such i must satisfy both $i \ge m+1$ and $\tau(k) = k$ for $k = i+1, \dots, n$. Since τ is injective we cannot have $\tau(i) > i$, hence $\tau(i) < i$. We have $j := \tau^{-1}(i) \ne i$ (since $\tau(i) \ne i$), and so $j < i$ by the choice of i as the largest non-fixed point of τ. Thus, we have $j < i$ but $\tau(j) = i > \tau(i)$. If $j \ge m$, it follows immediately from the definition of F that $\tau \notin F$, and hence $\tau \notin G$ *a fortiori* since $G \subset F$. On the other hand, if $j < m$, let $\sigma = (jm) \in S_m$ be the permutation that transposes j and m. Then we have $\tau\sigma(m) = \tau(j) = i > \tau(i) = \tau\sigma(i)$. Since $m < i$ but $\tau\sigma(m) > \tau\sigma(i)$, we have $\tau\sigma \notin F$ in this case, and so $\tau\sigma \notin G$. Since G is a group and $\sigma \in S_m \subset G$, it follows that $\tau \notin G$ (otherwise we would have $\tau\sigma \in G$). Thus, G contains no permutations $\tau \notin S_m$, so $G = S_m$.

Also solved by Paul Budney, Robert Calcaterra, William Cowieson, Dmitry Fleischman, Neville Fogarty, Abhay Goel, Tom Jager, Peter McPolin (Northern Ireland), Michael Reid, Nikhil Sahoo, and the proposer. There was 1 incomplete or incorrect solution.

A limit-ratio test for convergence to zero **June 2018**

2047. *Proposed by George Stoica, Saint John, New Brunswick, Canada.*

Let (a_n) be a sequence of nonzero real numbers such that

$$\lim_{n\to\infty} n\left(\left|\frac{a_n}{a_{n+1}}\right| - 1\right)$$

exists and is strictly positive. Prove or disprove: The sequence (a_n) is necessarily convergent.

Solution by Nikhil Sahoo (student), Berkeley City College, Berkeley, CA.

We show that the sequence (a_n) converges to zero under the weaker hypothesis that $L := \liminf_{n\to\infty} n(|a_n/a_{n+1}| - 1)$ is strictly positive.

Lemma. For any positive integers m, n, we have

$$0 < P_{m,n} := \prod_{k=1}^{n-1} \frac{mk}{mk+1} \le \frac{1}{\sqrt[m]{n}}.$$

(Per the usual convention on empty products, we let $P_{m,1} = 1$.)

Proof. Clearly, $P_{m,n} > 0$. The function $x \mapsto x/(x+1)$ is positive and increasing on $(0, \infty)$; therefore,

$$(P_{m,n})^m = \prod_{j=0}^{m-1} P_{m,n} = \prod_{k=1}^{n-1}\prod_{j=0}^{m-1} \frac{mk}{mk+1} \le \prod_{k=1}^{n-1}\prod_{j=0}^{m-1} \frac{mk+j}{mk+j+1} = \prod_{i=m}^{mn-1} \frac{i}{i+1} = \frac{m}{mn} = \frac{1}{n}.$$

Taking m-roots of both sides of the inequality above concludes the proof of the lemma.

For fixed M, it follows from the lemma that $\lim_{n\to\infty} P_{M,n} = 0$ since $1/\sqrt[M]{n} \to 0$ as $n \to \infty$. By the assumption that the limit in the statement of the problem is positive, there exist positive integers M and N such that $n \ge N$ implies $n(|a_n/a_{n+1}| - 1) \ge 1/M$; equivalently,

$$|a_{n+1}| \le \frac{Mn}{Mn+1} \cdot |a_n| \qquad \text{for all } n \ge N.$$

It follows by induction that $|a_n| \le |a_N| \cdot \prod_{k=N}^{n-1} \left[Mk/(Mk+1)\right] = |a_N| \cdot P_{M,n}/P_{M,N}$ for $n \ge N$. Since $\lim_{n\to\infty} P_{M,n} = 0$, we see that (a_n) converges to zero.

Editor's Note. Christopher Hammond remarked that the value of the limit $L = \lim_{n\to\infty} n(|a_n/a_{n+1}| - 1)$ is closely related to Raabe's test for convergence of the series $\sum_{n=1}^{\infty} a_n$. Christopher N. B. Hammond, The Case for Raabe's Test, *Mathematics Magazine* (forthcoming). The series converges absolutely when $L > 1$ and diverges when $L < 0$. For $0 \le L < 1$, the series may either diverge or converge conditionally (Theorem 1 therein). However, if $L > 0$ then the sequence (a_n) necessarily converges to zero as asserted in the statement of the problem (Proposition 2 therein).

Also solved by Ulrich Abel & Vitaliy Kushnirevych (Germany), Robert A. Agnew, Paul Budney, William Cowieson, Souvik Dey, Joseph DiMuro, Robert L. Doucette, Dmitry Fleischman, Abhay Goel, Russell Gordon, Christopher N. B. Hammond, Lixing Han, Eugene A. Herman, Tom Jager, John C. Kieffer, Jimin Kim (South Korea), Elias Lampakis (Greece), Kee-Wai Lau, Peter McPolin (Northern Ireland), Albert Natian, Northwestern University Math Problem Solving Group, Moubinool Omarjee (France), Michael Reid, Celia Schacht, Christopher Sinkule, Nora Thornber, Lawrence R. Weill, and the proposer. There were 2 incomplete or incorrect solutions.

A random triangle with vertices in a three-quarter disk June 2018

2048. *Proposed by Julien Sorel, Piatra Neamt, PNI, Romania.*

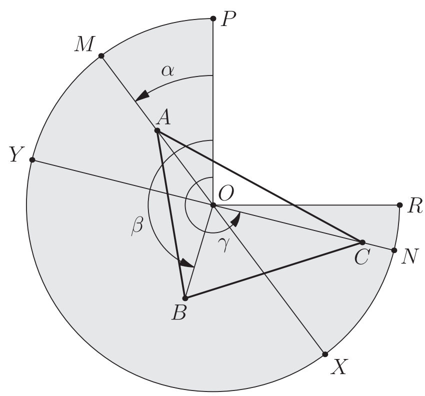
Three points A, B, C are chosen uniformly at random in the three-quarter disk

$$\mathcal{Q} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1, \text{ and either } x \le 0 \text{ or } y \le 0\}$$

obtained by removing the first quadrant of the unit disk. What is the probability that the origin $O = (0, 0)$ lies inside $\triangle ABC$?

Solution by Xueshi Gao (student), Peking University, Beijing, China.
We prove that the event $\mathcal{E}$ that O lies inside $\triangle ABC$ has probability $\mathbf{P}[\mathcal{E}] = 5/27$.

Let $P = (0, 1)$, $R = (1, 0)$, and $\alpha = \angle POA$, $\beta = \angle POB$, $\gamma = \angle POC$, so that $\alpha, \beta, \gamma \in [0, 3\pi/4]$. Let $\overline{MX}$ be the diameter through A so $AM < AX$, and $\overline{NY}$ the diameter through C so $CN < CY$. Consider the event $\mathcal{E}' = \mathcal{E} \cap \{\alpha < \beta < \gamma\}$ as shown in the figure below.



By the assumption that O lies inside $\triangle ABC$, each of the three angles $\angle COA$, $\angle AOB$ and $\angle BOC$ must be strictly less than a half revolution. It follows that point A must lie in the second quadrant, C in sector ROX, and B in sector XOY, that is,

$$0 \le \alpha < \frac{\pi}{2}, \quad \alpha + \pi < \gamma \le \frac{3}{2}\pi, \quad \text{and} \quad \gamma - \pi < \beta < \alpha + \pi.$$

Observe that angles α, β, γ are independent and uniformly distributed in $[0, 3\pi/2]$ because A, B, C are independent and uniformly distributed in $\mathcal{Q}$; therefore, the probability of event $\mathcal{E}'$ is

$$\mathbf{P}\left[\mathcal{E}'\right] = \int_0^{\frac{\pi}{2}} \int_{\alpha+\pi}^{\frac{3}{2}\pi} \int_{\gamma-\pi}^{\alpha+\pi} \left(\frac{2}{3\pi}\right)^3 d\beta\, d\gamma\, d\alpha = \frac{5}{162}.$$

By independence of the three random points A, B, C and the fact that event $\mathcal{E}$ is invariant under permutations thereof, we have

$$\mathbf{P}\left[\mathcal{E}\right] = \frac{\mathbf{P}\left[\mathcal{E}'\right]}{\mathbf{P}[\{\alpha < \beta < \gamma\}]} = \frac{5/162}{1/6} = \frac{5}{27}.$$

Also solved by Herb Bailey & Mark Bailey, Elton Bojaxhiu (Germany) & Enkel Hysnelaj (Australia), Robert Calcaterra, Bill Cowieson, Bao Do, Gregory Dresden, John N. Fitch, Neville Fogarty, Kyle Gatesman, GWstat Problem Solving Group, Elias Lampakis (Greece), Albert Natian, Mingyu Park (Korea), Sung Hee Park (Korea), Nikhil Sahoo, Jacob Siehler, Nora S. Thornber, Lawrence Weill, and the proposer. There were 2 incomplete or incorrect solutions.

Non-sparse disjoint subsets of any finite collection of squares **June 2018**

2049. *Proposed by Scott Duke Kominers, Harvard University, Cambridge, MA.*

Show that any finite set of squares in the plane (possibly of different sizes and not necessarily disjoint) has a subset consisting of non-overlapping squares that together cover at least 7% of the area covered by the full set.

Solution by Jimin Kim (student), Institute of Science Education for the Gifted and Talented, Yonsei University, Republic of Korea.
Let $\mathcal{C}$ be the given finite collection of squares. The assertion is trivial when $\mathcal{C}$ is empty, so we assume $\mathcal{C}$ is nonempty henceforth. Successively choose squares $S_1, S_2, \ldots$ in $\mathcal{C}$ by the following recursive method:

- Let S_1 be any square of largest area in $\mathcal{C}$.
- Having chosen $S_1, \ldots, S_i$, let S_{i+1} be any square in $\mathcal{C}$ having largest area among those squares in $\mathcal{C}$ that intersect neither of $S_1, \ldots, S_i$. If there is no such square, the procedure terminates.

The procedure must terminate after choosing a finite number $m \geq 1$ of squares $S_1, S_2, \ldots, S_m$ since $\mathcal{C}$ is finite by hypothesis; because of this, every square in $\mathcal{C}$ intersects some S_i. (Were there any squares in $\mathcal{C}$ intersecting no S_i, any such of largest area would allow the recursive procedure to continue!) Thus, we have $\mathcal{C} = \bigcup_{i=1}^{m} \mathcal{N}_i$ where $\mathcal{N}_i$ $(i = 1, \ldots, m)$ denotes the set of squares in $\mathcal{C}$ that intersect S_i but no S_j with $j < i$. Certainly, we have $S_i \in \mathcal{N}_i$ (since S_i intersects itself, but is otherwise chosen not to intersect S_j for any $j < i$). The sets $\mathcal{N}_1, \ldots, \mathcal{N}_m$ are clearly disjoint by construction; in particular, the squares $S_1, \ldots, S_m$ are disjoint. Moreover, S_i by choice has largest area among all squares in $\mathcal{N}_i \cup \cdots \cup \mathcal{N}_m$ (which is the set of squares in $\mathcal{C}$ intersecting none of $S_1, \ldots, S_{i-1}$); in particular, S_i has largest area among all squares in $\mathcal{N}_i$. We will show that the set consisting of the disjoint squares $S_1, \ldots, S_m$ covers at least 7% of the area covered by all squares in $\mathcal{C}$.

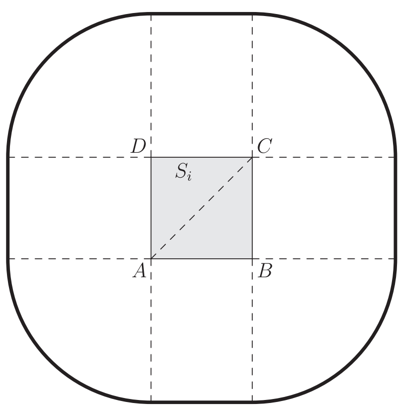


Figure 1 The region R_i consisting of points lying at distance no more than d from square $S_i = \square ABCD$. All dashed segments have the same length $d = AC$.

Let S_i be a square $\square ABCD$ with side $\ell = AB$ and diagonal $d = \sqrt{2}\ell = AC$. Any square $Q \in \mathcal{N}_i$ intersects S_i and has area, hence also diagonal length, not exceeding that of S_i. It follows that Q is fully covered by the region R_i (depicted in Figure 1 above) consisting of points at distance no more than d from S_i. Thus, the region $\overline{\mathcal{N}}_i$ covered by all the squares in $\mathcal{N}_i$ is fully covered by R_i, and its area $\left|\overline{\mathcal{N}}_i\right|$ satisfies

$$\left|\overline{\mathcal{N}}_i\right| \leq |R_i| = (2\pi + 4\sqrt{2} + 1)|S_i|.$$

(We denote by $\left|\mathcal{X}\right|$ the area of a region $\mathcal{X}$ of the plane.) The collection $\mathcal{C}$ covers the region $\overline{\mathcal{C}} = \overline{\mathcal{N}}_1 \cup \cdots \cup \overline{\mathcal{N}}_n$, so we have

$$\left|\overline{\mathcal{C}}\right| = \left|\bigcup_{i=1}^{m} \overline{\mathcal{N}_i}\right| \leq \sum_{i=1}^{m} \left|\overline{\mathcal{N}}_i\right| \leq (2\pi + 4\sqrt{2} + 1) \sum_{i=1}^{m} |S_i|.$$

Therefore, the disjoint squares $S_1, \ldots, S_m$ cover an area

$$\sum_{i=1}^{m} |S_i| \geq \frac{1}{2\pi + 4\sqrt{2} + 1} \left|\overline{\mathcal{C}}\right| = 0.077 \cdots \times \left|\overline{\mathcal{C}}\right|$$
$$> 0.07 \times \left|\overline{\mathcal{C}}\right|.$$

Also solved by Elton Bojaxhiu (Germany) & Enkel Hysnelaj (Australia), Robert Calcaterra, Kyle Gatesman, Abhay Goel, Sarah Kapinos & J. Todd Lee, Northwestern University Math Problem Solving Group, Nikhil Sahoo, Celia Schacht, Lawrence R. Weill, and the proposer.

Counting de Bruijn sequences of pairs of three symbols **June 2018**

2050. *Proposed by Sung Soo Kim, Hanyang University, Ansan, Korea.*

Find the number of sequences $a_1, a_2, \ldots, a_9$ in $\{1, 2, 3\}$ such that

(*i*) $a_1 = a_2 = 1$, and
(*ii*) the nine pairs $(a_1, a_2), (a_2, a_3), \ldots, (a_8, a_9), (a_9, a_1)$ are the same as the nine pairs $(1, 1), (1, 2), \ldots, (3, 2), (3, 3)$ in some order.

Solution by Skidmore College Problem Group, Saratoga Springs, NY.

Figure 2 Graph A (left) and graph B (right).

We show that there are 24 such sequences. This problem amounts to finding the number of cycles traversing each edge (i.e., Eulerian cycles) in the directed graph A on the left of Figure 2 above, beginning with the loop $(1, 1)$: The sequences $a_1 = 1, a_2 = 1, a_3, \dots, a_9$ of vertices (excluding the final return vertex $a_{10} = 1 = a_1$) in such a cycle are precisely those considered in the problem. We call these "*a-cycles*." There is a four-to-one correspondence between the set of a-cycles and the set of Eulerian cycles successively visiting vertices $1 = b_1, b_2, \dots, b_6$ of graph B on the right of Figure 2 above, starting at vertex 1 (and eventually traversing all edges, returning to vertex $b_7 = 1 = b_1$). We refer to the latter as "*b-cycles*." Indeed, given an a-cycle α we may simply remove the loops $(1, 1)$, $(2, 2)$, $(3, 3)$ from α to obtain a b-cycle β; conversely, given a b-cycle β, simply add the loop $(1, 1)$ at the beginning of β, insert the loop $(2, 2)$ at either of the two occasions when vertex 2 is visited, and the loop $(3, 3)$ at either of the two occasions when vertex 3 is visited—this gives four different a-cycles $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ the removal of whose loops results in β. We proceed to count the number of distinct b-cycles.

With graph B as depicted, we call directed edges $(1, 2)$, $(2, 3)$, and $(3, 1)$ *inner*, and edges $(2, 1)$, $(3, 2)$, and $(1, 3)$ *outer*. A moment's reflection shows that, given a vertex b_i of any path on graph B, the next vertex b_{i+1} is uniquely determined by specifying whether b_{i+1} is reached from b_i by traversing an inner or an outer edge. The set of b-cycles is partitioned into the disjoint sets of *inner* b-cycles that start with the inner edge $b_1 = 1 \to 2 = b_2$, and the set of *outer* b-cycles that start with the outer edge $b_1 = 1 \to 3 = b_2$. Given the identical role played by vertices 2 and 3 in the problem, it is clear that there are equally many inner and outer b-cycles. (Exchanging the labels of vertices 2 and 3 is a bijection between the sets of inner and outer b-cycles.) Thus, it suffices to count inner b-cycles. These are easily seen to be encoded by the sequences

iooooii, iioooi, iiiooo

(where "i" means "follow inner edge," while "o" means "follow outer edge"), corresponding to the b-cycles:

$$1 \to 2 \to 1 \to 3 \to 2 \to 3(\to 1),$$
$$1 \to 2 \to 3 \to 2 \to 1 \to 3(\to 1),$$
$$1 \to 2 \to 3 \to 1 \to 3 \to 2(\to 1).$$

Thus, there are 3 distinct inner b-cycles, while the number of b-cycles is $2 \cdot 3 = 6$, and the number of a-cycles (hence of sequences solving the problem) is $4 \cdot 6 = 24$.

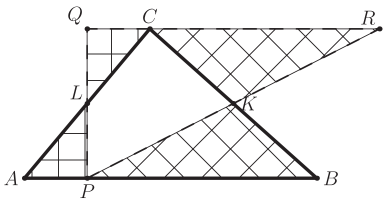
Editor's Note. Rob Pratt remarked that the problem asks for the number of de Bruijn sequences of order $n = 2$ over an alphabet of $k = 3$ symbols, of which there are $k!^{k^{n-1}}/k^n$ in general, and in particular $3!^{3^{2-1}}/3^2 = 24$ in this problem. (T. van Aardenne-Ehrenfest and N. G. de Bruijn, Circuits and trees in oriented linear graphs. *Simon Stevin* **28** (1951) 203–217.) Jacob Siehler brought to our attention that the twenty-four sequences and the general formula above appeared in this MAGAZINE. (Anthony Ralston, De Bruijn sequences—A model example of the interaction of discrete mathematics and computer science, *Mathematics Magazine* **55** (1982) 131–143.)

Also solved by Skyler Addy & Zachary Parker, Brian D. Beasley, Elton Bojaxhiu (Germany) & Enkel Hysnelaj (Australia), Robert Calcaterra, Timothy Crane, Dmitry Fleischman, Neville Fogarty, Abhay Goel, Eugene A. Herman, Dain Kim (Korea), Brad Meyer, Ioana Mihăilă, North Carolina Wesleyan College Fall 2018 MAT 318 Discrete Methods Class, Rob Pratt, Nikhil Sahoo, Joel Schlosberg, Jacob Siehler, David Stone & John Hawkins, Lawrence R. Weill, and the proposer. There was 1 incomplete or incorrect solution.

Answers

Solutions to the Quickies from pages 231 and 232.

A1091. Label the vertices of the given triangle $\triangle ABC$ so $\angle C$ is its largest angle; thus, $\angle A$ and $\angle B$ are both acute. Let K be the midpoint of $\overline{BC}$ and L the midpoint of $\overline{AC}$. Choose P on $\overline{AB}$ so $\overline{LP}$ is perpendicular to $\overline{AB}$. Note that P lies on side $\overline{AB}$ since $\angle A$ and $\angle B$ are both acute. Cut triangle $\triangle ABC$ along segments $\overline{KP}$ and $\overline{LP}$ thus splitting it into triangles $\triangle ALP$, $\triangle BKP$ and quadrilateral $CKPL$. Keeping the quadrilateral fixed, rotate the right triangle $\triangle ALP$ half a revolution about L to obtain a new triangle $\triangle CLQ$, and triangle $\triangle BKP$ half a revolution about K to obtain triangle $\triangle CKR$. In doing so we obtain a triangle $\triangle PQR$ with angle $\angle Q$ right. (This construction works even if the original triangle was a right triangle to begin.)



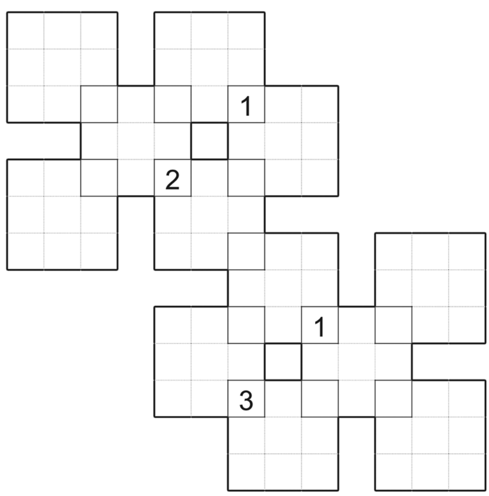
A1092. We show that the double integral has the value $\pi/4$ for all n. Swapping the order of integration and the names of the variables x, y, we have

$$I = \int_0^1 \int_0^1 \arctan\left(\frac{x^n}{y^n}\right) dx\, dy = \int_0^1 \int_0^1 \arctan\left(\frac{y^n}{x^n}\right) dx\, dy.$$

From the identity $\arctan(t) + \arctan(1/t) = \pi/2$ (valid for $t > 0$), we obtain

$$I = \frac{1}{2} \int_0^1 \int_0^1 \left[\arctan\left(\frac{x^n}{y^n}\right) + \arctan\left(\frac{y^n}{x^n}\right)\right] dx\, dy = \frac{1}{2} \int_0^1 \int_0^1 \frac{\pi}{2}\, dx\, dy = \frac{\pi}{4}.$$

TRIBUS Puzzle



How to play. Fill each of the three-by-three squares with either a 1, 2, or 3 so that each number appears exactly once in each column and row. Some cells apply to more than one square, as the squares overlap. Each of the three-by-three squares must be distinct. The solution can be found in page 172.

— David Nacin, William Paterson University, Wayne, NJ (nacind@wpunj.edu)

REVIEWS

PAUL J. CAMPBELL, *Editor*
Beloit College

Assistant Editor: Eric S. Rosenthal, West Orange, NJ. Articles, books, and other materials are selected for this section to call attention to interesting mathematical exposition that occurs outside the mainstream of mathematics literature. Readers are invited to suggest items for review to the editors.

Pitici, Mircea, *The Best Writing on Mathematics 2018*, Princeton University Press, 2019; xvii + 250 pp, $24.95. ISBN 978-0-691-18276-6.

This new annual collection reprints Francis Edward Su's passionate "Mathematics for human flourishing" ("mathematics cultivates virtues that help people flourish"); offers a reflection by Robert S.D. Thomas on the importance of "being interesting" as an aesthetic value in mathematics (unfortunately, a little over-written in hesitant philosophy-speak); and includes a definition of computational thinking by Peter J. Denning (who urges that computer scientists learn something about the field where their computation is to be used). Among the 15 other essays is an explanation by Arthur Benjamin et al. of the "bingo paradox," that the winning card very likely has a horizontal row rather than a vertical column.

Tijms, Henk, *Surprises in Probability: Seventeen Short Stories*, CRC Press, 2019; xii + 131 pp, $99.95, $39.95 (P). ISBN 978-0-367-00082-0, 978-0-367-00043-1.

These are not short stories in the literary sense but delightful vignettes about probability, with many real-life examples of the gambler's ruin, birthday coincidences, Bayes' rule, the Benford distribution, the Kruskal count, lotto, the Kelly investment criterion, optimal stopping, and more. But the book's prices are shameful.

Zimmerman, Paul, and 15 co-authors, *Computational Mathematics with SageMath*, SIAM, 2018; xiv + 464 pp, $69 (P). ISBN 978-1-611975-45-1.

Budgetary considerations at my institution recently threatened (briefly) our institutional license of the expensive computer algebra system Mathematica. Other commercial alternatives are also expensive. What would I have done? Well, the open-source Sage system is free. The first quarter of this book is a manual on features of Sage, but the rest shows specifically how to do all kinds of mathematics: number theory, finite fields, polynomial rings, linear algebra, difference and differential equations, numerical computation, combinatorics, graph theory, and linear programming. Each chapter has exercises, with answers provided in an appendix.

Zagare, Frank C., *Game Theory, Diplomatic History and Security Studies*, Oxford University Press, 2019; xv + 185 pp, $95, $45.95 (P). ISBN 978-0-19-883158-7, 978-0-19-883159-4.

"The main purpose of this book is to demonstrate, by way of example, the several advantages of using a formal game-theoretic framework to explain complex [historical] events and relationships." An initial chapter gives a brief introduction to game theory; subsequent chapters apply the theory to the Moroccan crisis of 1905–6 (an incident leading to World War I), the Cuban missile crisis of 1962 (metagames and the theory of moves come into play, with a further perspective on contemporary relations between the U.S. and North Korea), and wider issues about the causes of World War I. Further chapters comment on the value of deterrence and speculate on reasons for the "long peace" of the Cold War.

Math. Mag. **92** (2019) 239–240. doi:10.1080/0025570X.2019.1603523.

Ording, Philip, *99 Variations on a Proof*, Princeton University Press, 2019; xi + 260 pp, \$19.95. ISBN 978-0-691-15883-9.

Wood, Dick A., 36 methods of mathematical proof, *Mathematics Teacher* 91 (8) (November 1998) 649.

To my mind, an ideal way to learn creative writing would be to try to write in the style of each of several particular authors. Author Ording has done exactly that, as an exercise in mathematics as a "literary or aesthetic medium." Inspired by Raymond Queneau's 1947 *Exercises in Style*, which retells a story in 99 ways, he proves a theorem in 99 ways. The theorem is "trivial," understandable to a student in high school algebra; the proofs range from "back of the envelope" to a screenplay featuring Cardano and Tartaglia, from a blog to American Sign Language. Some of them ("Omitted with condescension," "Authority") hark back to Dick Wood's "36 methods...": "by obviousness," "by intimidation," "by tautology," etc. Spoiler ahead!: The theorem is: If $x^3 - 6x^2 + 11x - 6 = 2x - 2$, then $x = 1$ or $x = 4$.

Singmaster, David, *Problems for Metagrobologists: A Collection of Puzzles with Real Mathematical, Logical or Scientific Content*, World Scientific, 2016; xi + 234 pp, \$58, \$28 (P). ISBN 978-981466363-2, 978-981466364-9.

Many mathematicians have contributed solutions to puzzles and problems in the problems sections of mathematical publications. Far fewer have contributed such puzzles and problems. Author Singmaster is among the latter, and this book is a collection of his puzzle problems from more than 50 years. The 221 puzzles are grouped by type: arithmetic, monetary, diophantine, alphametics, logic, geometrical, geographic, calendrical, physical, combinatorial, and more. Of course, solutions are given, together with, in some cases, the source or inspiration. I was surprised to learn that the new-to-me word "metagrobologist" is not a neologism coined by Singmaster, but that it derives from a French verb used by Rabelais 500 years ago to mean "to puzzle, mystify," and the English version was used by Kipling. Now if only Singmaster can complete his voluminous bibliography of sources in recreational mathematics! (See https://www.puzzlemuseum.com/singma/singma-index.htm for a preliminary version.)

Yau, Shing-Tung, and Steve Nadis, *The Shape of a Life: One Mathematician's Search for the Universe's Hidden Geometry*, Yale University Press, 2019; xvi + 293 pp, \$28. ISBN 978-0-300-23590-6.

Author Yau, who was awarded a Fields Medal in 1983 and other significant honors, tells in great detail the story of his life, and more importantly, his interpretation in personal terms of its events. The book's subtitle is a bit misleading; while Yau's research may have been such a search, the book is about his route from China, via Hong Kong, to UC-Berkeley and a career in the U.S., plus his many returns to China to found mathematical institutes. Mathematics majors from China may find this book particularly inspiring, especially in the gratitude that Yau expresses to his family and many colleagues.

Getty, Daniel, Hao Li, Masayuki Yano, Charles Gao, and A.E. Hosoi, Luck and the law: Quantifying chance in fantasy sports and other contests, *SIAM Review* 60 (4) (December 2018) 869–887.

Some U.S. states and communities outlaw gambling of various kinds, including "betting" on particular games, including office pools and other contests with cash prizes. Notorious past examples (which have led to litigation) include bingo and poker. A contemporary instance is fantasy sports competitions, "games in which participants assemble virtual teams of athletes and compete [for cash] based on the athletes' real-world statistical performance." At the heart of the legal matter is whether a game is predominantly a game of chance (hence gambling) or or a game of skill (not gambling). This article's authors skirt the legal aspect entirely and concentrate on analyzing data from fantasy sports, comparing the outcomes of real players with imaginary players playing randomly. They define a measure of relative skill and apply it not only to fantasy sports but also to real sports, coin flipping, and mutual funds. The measure for each fantasy sport (hockey, baseball, football, basketball) is close to that for its real counterpart, with all showing a significant role for skill. I will not spoil your day by relating where mutual funds stand on their measure.