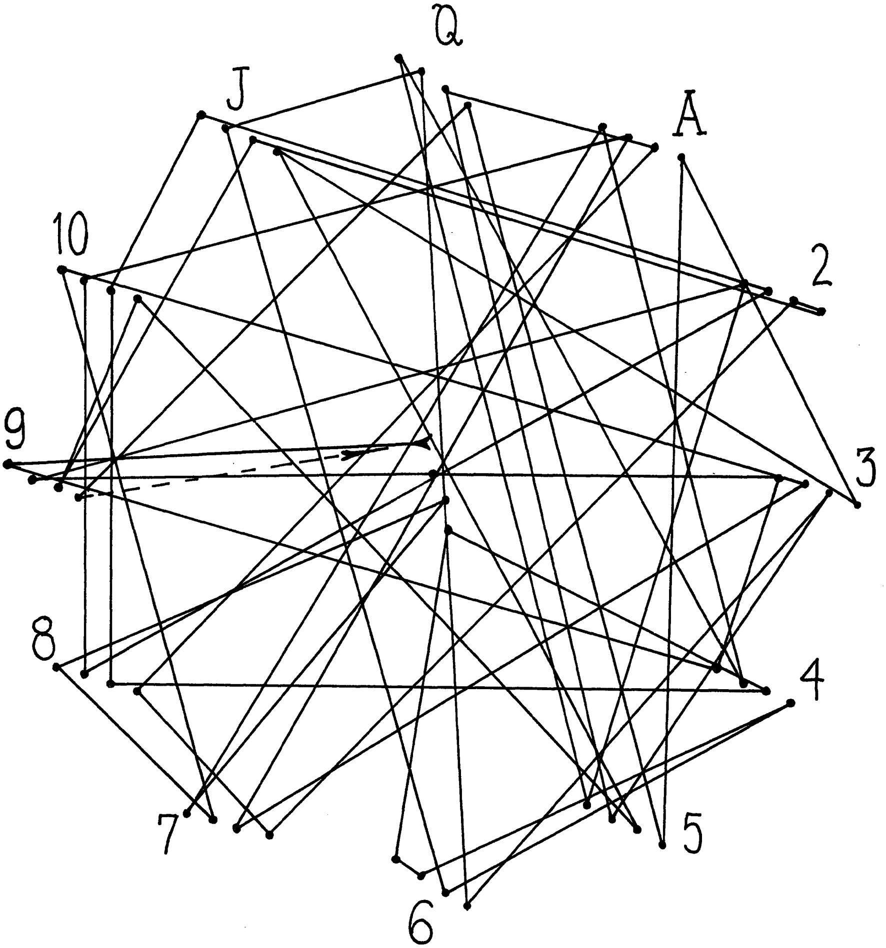
MATHEMATICS MAGAZINE



Vol. 54, No. 4
September, 1981

BOOLE'S ALGEBRA • CLOCK SOLITAIRE ANALYSED
CONTINUED FRACTIONS • CASTING OUT NINES

MATHEMATICS MAGAZINE

Volume 54, Number 4, September 1981
(ISSN—0025-570X)

COVER: Lines connect successive positions drawn in a winning game of Clock Solitaire (p. 202). Design by the Editor.

ARTICLES

NOTES

PROBLEMS

REVIEWS

NEWS AND LETTERS

EDITORIAL POLICY

Mathematics Magazine is a journal of collegiate mathematics which aims to provide inviting, informal mathematical exposition of interest to undergraduate students. Manuscripts accepted for publication in the *Magazine* should be written in a clear and lively expository style and stocked with appropriate examples and graphics. Our advice to authors is: say something new in an appealing way or say something old in a refreshing way. The *Magazine* is not a research journal and so the style, quality, and level of articles submitted for publication should realistically permit their use to supplement undergraduate courses. The editor invites manuscripts that provide insight into the history and application of mathematics, that point out interrelationships between several branches of mathematics and that illustrate the fun of doing mathematics.

New manuscripts should be sent to: Doris Schattschneider, Editor, Mathematics Magazine, Moravian College, Bethlehem, Pennsylvania 18018. Manuscripts should be prepared in a style consistent with the format of *Mathematics Magazine*. They should be typewritten and double spaced on $8\frac{1}{2}$ by 11 paper. Authors should submit the original and one copy and keep one copy as protection against possible loss. Illustrations should be carefully prepared on separate sheets of paper in black ink, the original without lettering and two copies with lettering added.

Authors planning to submit manuscripts should read the full statement of editorial policy which appears in the News and Letters section of this *Magazine*, Vol. 54, pp. 44–45. Additional copies of the policy are available from the editor.

AUTHORS

Ian Richards ("Continued Fractions Without Tears") is a professor of mathematics at the University of Minnesota. He attended college there, did his graduate work at Harvard, taught for two years at MIT, and then returned to Minnesota. His field of research is analysis, where he has worked in the areas of complex variables, real variables, analytic number theory, and recursive analysis. His hobbies include whitewater and wilderness canoeing and reading poetry (about which he knows little but claims to have good taste).

Theodore Hailperin ("Boole's Algebra Isn't Boolean Algebra"), professor of mathematics at Lehigh University, has just retired from teaching. His special interests are in the fields of mathematical logic and foundations of mathematics. He has directed NSF studies concerned with mathematical techniques used to study new kinds of reasoning and theorem-proving with computers. He is author of *Boole's Logic and Probability*, North-Holland, 1976.

ILLUSTRATIONS

Vic Norton, Bowling Green State University, illustrated the scene on p. 171.

Gerald Porter, U. of Pennsylvania, produced the computer-drawn graphs for "On Functions Whose Inverse Is Their Reciprocal."

The cartoon (p. 207) by **Bil Keane** is from "The Family Circus," reprinted courtesy of the Register and Tribune Syndicate, Inc.

All other illustrations were provided by the authors.

BUSINESS INFORMATION. Mathematics Magazine is published by the Mathematical Association of America at Washington, D.C., five times a year in January, March, May, September, and November. Ordinary subscriptions are $18 per year. Members of the Mathematical Association of America or of Mu Alpha Theta may subscribe at special reduced rates. College and university mathematics departments may purchase bulk subscriptions for distribution to undergraduate students. A bulk subscription consists of five or more copies of each issue to a single address for one year; the annual bulk subscription rate is $5.00 per copy in the package, with a minimum $25.

Subscription correspondence and notice of change of address should be sent to A. B. Willcox, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, D.C. 20036. Back issues may be purchased, when in print, from P. and H. Bliss Co., Middletown, Connecticut 06457.

Advertising correspondence should be addressed to Raoul Hailpern, Mathematical Association of America, SUNY at Buffalo, Buffalo, New York 14214.

Second class postage paid at Washington, D.C., and additional mailing offices.

Continued Fractions Without Tears

Ping-pong using Farey sequences is proposed as an alternative to the traditional fraction chain.

IAN RICHARDS
University of Minnesota
Minneapolis, MN 55455

The traditional presentation of continued fractions is via an infinite sequence of quotients within quotients. For example:

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cdots}}}.$$

This "python descending a staircase" format has the advantage of historical priority, and it suggests important generalizations. However, there are approaches that are more conceptual, and I would like to outline one of them. (For the sake of comparison, we will untangle the infinite fraction in its traditional form and tie it together with our theory at the end of the paper.) If by the end of this paper you feel that this is all pretty trivial, then I have succeeded; if you think it's tough, then this approach isn't your cup of tea. Our goal will be to discover and prove the essential facts about continued fractions and to develop a computer algorithm for them.

The spirit of this presentation is geometrical, but it is geometry in an unusual setting: that of one-dimensional space. Remember the traditional drill sergeant's complaint: "Can't you tell your left hand from your right hand?" Our approach will be, essentially, to keep careful track of which points lie right and which ones lie left.

The Farey process

Now let's take a look at the theory. A pair of nonnegative fractions,

$$\frac{a}{b} < \frac{c}{d},$$

is called a **Farey pair** if $bc - ad = 1$. This means, of course, that the difference between the fractions is $1/bd$. The **mediant** of these two fractions is defined to be $(a+c)/(b+d)$. For example, $2/3 < 3/4$ is a Farey pair, with mediant $5/7$. A trivial calculation shows that the mediant always lies between a/b and c/d, so we have:

$$\frac{a}{b} < \underset{\text{(mediant)}}{\frac{a+c}{b+d}} < \frac{c}{d}.$$

For convenience, we call the interval $[a/b, c/d]$ a **Farey interval** if a/b and c/d are a Farey pair.

These ideas have a curious history. Farey discovered some properties of the mediant but was unable to prove them; the proof was supplied by Cauchy, who named the theory after its supposed discoverer. However, both were unaware that Haros had proved the same theorems several years before. A poignant comment on the capriciousness of fame was made by Hardy and Wright [**2**]: "Farey has a notice of twenty lines in the [British] 'Dictionary of National Biography,' where he is described as a geologist. As a geologist he is forgotten, and his biographer does not mention the one thing in his life which survives."

LEMMA 1. *Let* $[a/b, c/d]$ *be a Farey interval and consider the mediant* $(a+c)/(b+d)$ (*see* FIGURE 1). *Then*:

(i) *the two subintervals formed by inserting the mediant are also Farey intervals*;

(ii) *among all the fractions* x/y *lying strictly between* a/b *and* c/d, *the mediant is the one* (*and only one*) *with the smallest denominator.*

As we shall see, part (ii) of Lemma 1 is the key to the whole theory.

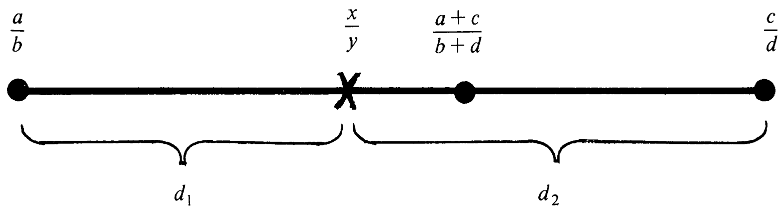


FIGURE 1. **The geometric setup in Lemma 1.**

Proof. Part (i) is easy. For part (ii), take any $x/y \neq (a+c)/(b+d)$ in the open interval $(a/b, c/d)$. We first assume that $y < (b+d)$ and derive a contradiction. The distance $d_1 = (x/y) - (a/b)$ (see FIGURE 1) is equal to $(bx - ay)/by$. The numerator $(bx - ay)$ is a positive integer and thus $\geqslant 1$; hence

$$d_1 \geqslant \frac{1}{by}.$$

Similarly

$$d_2 = \frac{c}{d} - \frac{x}{y} \geqslant \frac{1}{dy},$$

and so

$$d_1 + d_2 \geqslant \frac{1}{by} + \frac{1}{dy} = \frac{b+d}{y} \cdot \frac{1}{bd}.$$

But since $a/b, c/d$ is a Farey pair, the distance $d_1 + d_2 = 1/bd$, and the assumption that $y < (b+d)$ leads to a contradiction.

Now we come to the case $y = b+d$. We can handle this without computation, using the following device. From part (i), the mediant $(a+c)/(b+d)$ partitions the original interval into two Farey subintervals. We apply what we already know to these subintervals. The mediants for these subintervals must have denominators larger than $(b+d)$. So by what we have already proved, there is no x/y *inside* either of these subintervals with $y \leqslant (b+d)$. (Yes, $\leqslant$.)

The procedure used in the last paragraph of the proof above provides an introduction to a technique which is called the **slow continued fraction algorithm**. This is a method for finding the "best" rational approximations to an irrational number α. (For convenience, we will assume that $0 < \alpha < 1$.) However, it turns out to be easier to forget about α for a moment, and study the technique in the absence of its object. In that case, the technique is called the **Farey process**. Later it will become clear how we apply the Farey process to zero in on α.

$\frac{0}{1}$								$\frac{1}{1}$
$\frac{0}{1}$				$\frac{1}{2}$				$\frac{1}{1}$
$\frac{0}{1}$		$\frac{1}{3}$		$\frac{1}{2}$		$\frac{2}{3}$		$\frac{1}{1}$
$\frac{0}{1}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{2}{5}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{1}{1}$

TABLE 1. **The Farey process.**

$\frac{1}{3}$								$\frac{2}{5}$
$\frac{1}{3}$				$\frac{3}{8}$				$\frac{2}{5}$
$\frac{1}{3}$		$\frac{4}{11}$		$\frac{3}{8}$		$\frac{5}{13}$		$\frac{2}{5}$
$\frac{1}{3}$	$\frac{5}{14}$	$\frac{4}{11}$	$\frac{7}{19}$	$\frac{3}{8}$	$\frac{8}{21}$	$\frac{5}{13}$	$\frac{7}{18}$	$\frac{2}{5}$

TABLE 2. **A continuation of Table 1, starting with the Farey pair 1/3 and 2/5.**

Here is the Farey process. Start with a Farey pair a/b and c/d. Take their mediant; this creates two subintervals and two new Farey pairs. Form all possible mediants again; this gives four subintervals.... The process is illustrated in TABLES 1 and 2. (For a curious property of this algorithm, see the paper of Shrader-Frechette in this *Magazine* [**3**].) Now it is clear how we zero in on an irrational number α. We simply, at each stage, keep the interval that contains α and discard the rest. This is the slow continued fraction algorithm. (The "fast" or standard algorithm is a refinement of the slow one, and we will describe it presently.)

If α were a rational number, $\alpha = p/q$, then the situation could be different, for p/q might appear as one of the division points in the Farey process. In fact that always happens! This, essentially, is the theorem that Farey discovered but couldn't prove. (Actually Farey looked at it differently, which is one reason why he couldn't prove it.) Farey's formulation—which has many uses, but is the wrong approach here—can be found in any standard textbook on number theory, e.g., [**2**], p. 23.

From now on, we will assume that the Farey process begins with the numbers $0/1$ and $1/1$.

THEOREM 1. *Every rational number p/q in lowest terms, with $0 < p/q < 1$, appears at some stage of the Farey process.*

EXAMPLE. The reader might want to continue the process in TABLE 2 until he finds the fraction $37/100$, which lies between $7/19$ and $3/8$.

Proof of Theorem 1. As expected, we use Lemma 1 above. After that, the proof almost writes itself. Suppose that a given fraction p/q between 0 and 1 never shows up in the Farey progression. Then at every stage, p/q remains squeezed between two adjacent fractions in the Farey process. By part (i) of Lemma 1, these fractions will always be a Farey pair. But the denominators of these fractions increase without bound (on one side of p/q, at least). Eventually the sum of the denominators exceeds q, and this violates Lemma 1, part (ii).

We seem to have strayed from our objective, which is the good approximation of irrational numbers by rationals. Now we come back to it. First we define what we mean by a "good" approximation.

DEFINITION 1. Let α be an irrational number with $0 < \alpha < 1$. Then a fraction p/q is called a **best left** (respectively, **best right**) approximation to α if:
(i) $p/q < \alpha$ (respectively, $p/q > \alpha$);
(ii) there is no fraction x/y between p/q and α with a denominator $y \leq q$.

Thus we put the left and right approximations to α into separate categories which do not compete against each other (like the American League and National League in baseball, before the World Series). After that, we give preference to fractions with small denominators. The small denominators, of course, are the whole point. (It doesn't take all this fuss to prove that the rational numbers are dense in the reals!) What we seek are classy approximations, like the famous estimates $22/7$ and $355/113$ for π. Thus $355/113$ gives π correctly to six decimal places, although the denominator in this fraction is scarcely over a hundred. [The next term in the continued fraction expansion for π cannot be found on a hand calculator, because calculators round off after about ten digits, and the next term is more accurate than that.] The following theorem shows how to discover these good approximations.

THEOREM 2. *Take any irrational number α, with $0 < \alpha < 1$. The slow continued fraction algorithm (= the Farey process, zeroed in on α) gives a sequence of best left and right approximations to α. Every best left/right approximation arises in this way.*

Proof. We use Lemma 1 and Theorem 1. Thus consider any Farey pair a/b and c/d. Lemma 1 tells us that all fractions lying between a/b and c/d have denominators larger than either b or d. Hence these fractions are not in competition with a/b or c/d, and automatically, a/b and c/d furnish best left/right approximations to all irrational numbers α lying between them. This proves the first part.

Now from Theorem 1, which tells us that all fractions occur in the Farey process, we will prove the second part. Take any fraction p/q between 0 and 1 and consider the first time it appears in the Farey process. Then p/q must be the mediant of its two neighbors: call them a/b and c/d. Thus we have the three adjacent terms (see also TABLE 2):

$$\frac{a}{b} < \frac{p}{q} < \frac{c}{d}$$

with $p = a + c, q = b + d$. Now there are two possibilities. If α lies between a/b and c/d (and hence in an interval also bounded by p/q), then p/q is a term in the slow continued fraction algorithm for α. If not (so that α lies left of a/b or right of c/d), then one of the fractions $a/b, c/d$ beats p/q on all counts, by being closer to α and having a smaller denominator; and thus p/q is not a "best" approximation to α.

The fast continued fraction algorithm

We turn now to the fast or standard continued fraction algorithm. This involves a selection of certain exceptionally good approximations furnished by the slow algorithm. It is important to note that the fast algorithm finds nothing new; in fact, it finds less, but by doing so goes faster. Here is a description of it.

Recall that the slow algorithm involves a series of shrinking Farey intervals zeroing in on an irrational number α. At each stage in this shrinking process, one end of the Farey interval moves in closer to α, and the other end stays put. Now it may happen that the same end, say the left one, moves several times in a row before the right-hand end moves again. Suppose that the left endpoint moves altogether s times. In the slow continued fraction algorithm we would keep all of the resulting points, since all of them furnish best left approximations to α. In the fast algorithm, we retain only the last (or sth) point.

To understand this better, it may be helpful to consult FIGURE 2. The left-hand endpoint $a/b = a_0/b_0$ moves successively to $a_1/b_1, a_2/b_2, \ldots, a_s/b_s$ and then stops. How are these a_k/b_k computed? Just ask, what is the Farey process? Each a_{k+1}/b_{k+1} is simply the mediant of a_k/b_k and c/d, and the algorithm successively computes these mediants until the mediant a_{s+1}/b_{s+1} is to the right of α. We stop the rightward migration of the left endpoints of the Farey pairs with the fraction a_s/b_s, before we cross the point α. This rightward migration is then followed by a similar leftward migration (of the right-hand endpoint) and so back and forth ad infinitum. It is conventional to index these migrations by n, the rightward moving ones being even, the others odd. I have suppressed the variable n to avoid double superscripts, but the reader should understand it is implicit. It is important to note that c/d was the fraction retained from the leftward migration immediately preceding the situation shown in FIGURE 2.

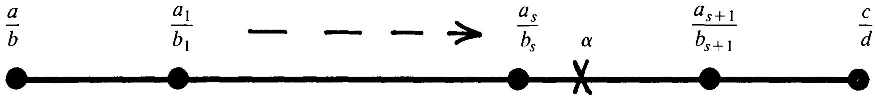


FIGURE 2. A stage in the slow continued fraction algorithm, in which the left endpoint a/b moves rightward s times. The next point a_{s+1}/b_{s+1} is on the wrong side of α. For each k, a_{k+1}/b_{k+1} is the mediant of a_k/b_k and c/d. The fast algorithm retains the fraction a_s/b_s and (when $s > 1$) discards all of the a_k/b_k with $1 \leq k < s$.

EXAMPLE. Referring to TABLE 2, let $a/b = 1/3$ and $c/d = 2/5$, so that c/d was the last term in a previous migration. Suppose that α lies between $5/13$ and $7/18$. Then the left endpoint $a/b = 1/3$ moves over two times, to $3/8$ and $5/13$, giving $s = 2$. The next rightward movement would bring the left endpoint to $7/18$, which is on the other side of α.

We have explained what the fast algorithm is. Now we ask: is it a good method? Have our choices of which points to keep, which discard, been good ones? We will define a precise sense in which the approximations which we have kept are better than those we threw away.

DEFINITION 2. Let α be any number with $0 < \alpha < 1$, and let p/q be a fraction. We define the **ultra-distance** from p/q to α to be $q|(p/q) - \alpha|$, i.e., q times the ordinary distance. We call p/q an **ultra-close approximation** to α if, among all fractions x/y with denominators $y \leq q$, p/q has the least ultra-distance to α. (Here we make no distinction between left and right approximations.)

Thus each fraction p/q is handicapped by having its distance to α multiplied by q. It is worth noting that between two fractions a/b and c/d, the ultra-distance is not commutative, because the denominators b and d are probably different. However, for our proof it is much more important (crucial, in fact) to look at the good side. The ultra-distance from p/q to α has nothing to do with the number theoretic structure of α. If α moves closer to p/q, then the ultra-distance from p/q to α decreases.

This may seem very artificial, and it would be nice to have an intrinsic idea of what the ultra-distance means. Luckily there is a very natural one. In what we have done so far it is really the denominators of fractions which are interesting. Suppose we consider together all of the fractions $0/q, 1/q, 2/q, \ldots$ with a fixed denominator q. A little thought shows that any number α can be approximated by one of these fractions to within a distance of $1/2q$. This approximation is guaranteed, so to speak. Now the ultra-distance is just the ratio

$$\frac{|(p/q) - \alpha|}{(1/q)}$$

Thus (ignoring the $1/2$) the ultra-distance tells us how much better the fraction p/q does than what we could automatically expect.

A comparison of Definitions 1 and 2 shows that "ultra-close" implies "best." (Just as with modern advertising, "best" isn't really very good.) Hence, by Theorem 2, the slow continued fraction algorithm furnishes the only possible candidates for "ultra-close" status. We aim to prove that the fast algorithm makes the correct choices from this list.

LEMMA 2. *Let a/b and c/d be a Farey pair with mediant $(a+c)/(b+d)$. Then the ultra-distances from either a/b or c/d to the mediant are the same.*

In terms of the ordinary distance on the number line, the mediant is not equidistant from a/b and c/d, but in terms of the ultra-distance, it is.

Proof. The ultra-distance from a/b to $(a+c)/(b+d)$ is

$$b\cdot\left(\frac{a+c}{b+d}-\frac{a}{b}\right),$$

and a simple calculation (using $bc-ad=1$) reduces this to $1/(b+d)$. By symmetry, the same holds for c/d.

THEOREM 3. *Take any irrational number* α, $0<\alpha<1$. *The fast continued fraction algorithm gives precisely the set of all ultra-close approximations to* α.

Proof. Since "ultra-close" approximations are also "best," it follows from Theorem 2 that we need consider only the terms produced by the slow continued fraction algorithm. Recall that the slow algorithm gave us a sequence of terms $a_1/b_1,\ldots,a_s/b_s$ moving inward towards α, and that we chose a_s/b_s and discarded the rest (see FIGURE 2). The question is: did we choose correctly; i.e., is a_s/b_s ultra-close to α? (Here we must not jump to conclusions. It is not a priori clear that the fractions which are closer to α are better, because the ultra-distance also involves the denominators, and the denominators in the sequence $a_1/b_1,\ldots,a_s/b_s$ are increasing.)

We should also remember that the slow algorithm involves a *sequence* of leftward and rightward migrations, and FIGURE 2 shows only a single stage in this process. By induction, we can assume that our theorem holds for all of the previous stages. In particular, we can assume that c/d (which was the term retained from the previous migration) is an ultra-close approximation to α. So now the question reduces to: which of the a_k/b_k (if any) are better than c/d?

Take any a_k/b_k. Recall that by definition of the Farey process, a_{k+1}/b_{k+1} is the mediant of a_k/b_k and c/d. Hence by Lemma 2, a_k/b_k and c/d have the same ultra-distance to a_{k+1}/b_{k+1}. Thus the contest is decided by whether α lies to the left or right of a_{k+1}/b_{k+1}. If left, a_k/b_k wins. If right, c/d wins. But (cf. FIGURE 2), α lies to the left of a_{k+1}/b_{k+1} $(1\leqslant k\leqslant s)$ if and only if $k=s$.

So the proof is, after all, just a matter of knowing your left hand from your right hand.

How to program it

The Farey process we have described is a concrete algorithm for approximation of an irrational number, and is suitable for programming on a computer or programmable calculator. Here, for the sake of brevity, we will give the reasoning as a chain of assertions, which the reader is invited to prove. We will conclude with an actual program. (We are not here concerned with the minutiae of programming languages, and common sense dictates that we continue to follow the notations of this paper.)

We assume that the fractions a/b and c/d in FIGURE 2 have already been found. Of course, we view this as a set of four integers a,b,c,d, not two real numbers. Our objective is to compute the integers a_s and b_s. To achieve this, we will first show how to compute a_k and b_k for any k, and then see how to find the stopping index s.

1. The integers a_k and b_k are given by $a_k=a+kc, b_k=b+kd$. (Hint: a_{k+1}/b_{k+1} is the mediant of a_k/b_k and c/d.)

2. The function $f(x)=(a+xc)/(b+xd)$ is strictly increasing for $x>0$. If we define the real number γ by the condition $f(\gamma)=\alpha$, then $\gamma=(\alpha b-a)/(c-\alpha d)$. (Note: this depends on the assumption that $a/b<c/d$; if that inequality were reversed, then we should replace the word "increasing" by "decreasing." We mention in passing that $(\alpha b-a)$ is the ultra-distance from a/b to α, and similarly for $(c-\alpha d)$.)

3. The stopping index s is the greatest integer $\leqslant\gamma$. (Hint: use the fact that $f(x)$ is increasing.) This data gives us the basis for a workable program. To make a "do loop" out of it, we compute γ, then s, then a_s and b_s. Then we make the replacements $a=c, b=d, c=a_s, d=b_s$ and start over.

4. There is a way to speed things up. Let γ and s be as above, and let γ' be the next value of γ. Then $\gamma'=1/(\gamma-s)$.

5. The best starting point for this program is $a/b = 0/1, c/d = 1/0$ which we think of as "infinity." Then after step 2, $\gamma = \alpha$. (By starting with $c/d = 1/0$ instead of $1/1$, we remove the restriction that $\alpha < 1$.)

Now let's write the program. The variables are $\gamma, s, a, b, c, d, a_s, b_s$, of which only γ is not a positive integer. The number to be approximated is α. The fractions a_s/b_s are the approximations.

PROGRAM.

Start with $a = 0, b = 1, c = 1, d = 0, \gamma = \alpha$.

Main Loop:

$s = \text{int}(\gamma)$

$$a_s = a + sc \quad (1)$$

$$b_s = b + sd \quad (2)$$

Print out a_s, b_s (and perhaps s)

$a = c$

$b = d$

$c = a_s$

$d = b_s$

$$\gamma = 1/(\gamma - s) \quad (3)$$

Repeat the main loop.

This is an infinite loop, of course; you can just stop the program by hand or else fix it. The variable s occurs in the standard presentations of continued fractions, where it is usually written a_n. If you take the number $e \cong 2.718\ldots$ you will find that the s values follow an interesting pattern; you will also find that this pattern breaks down eventually (why?). The s values for the square roots of various integers are quite interesting. Try also the golden mean $(1+\sqrt{5})/2$; here the b_s values form a well-known sequence, and the s values are also interesting. Nothing is known about higher level algebraic irrationals. A famous unsolved problem is to prove that the sequence of s values is unbounded for any algebraic number of degree greater than two.

If we apply our program to the number π, we obtain the following data:

a_s	b_s	s
3	1	3
22	7	7
333	106	15
355	113	1
103,993	33,102	292
.	.	.
.	.	.
.	.	.

There is little point in carrying the calculations much further, since the approximation $a_s/b_s = 103{,}993/33{,}102$ is already good to nine decimal places, and machine round-off error would soon render the results meaningless. (For example, the same program on my hand calculator produced a 293 in place of the 292 at the bottom of the s column.) Looking at some of the other fractions

a_s/b_s, we find the familiar approximations 22/7 and 355/113 mentioned earlier. Finally we observe that the numbers in the s column are just those which appeared in the "python" at the beginning of the article. This is no accident, as we now show.

Relation to the traditional format

Following our notations, the continued fraction expansion of an irrational number $\alpha > 0$ can be expressed in closed form by the equation:

$$\alpha = s_0 + \cfrac{1}{s_1 + \cfrac{1}{s_2 + \cfrac{\ddots}{\ddots + \cfrac{1}{s_{n-1} + \cfrac{1}{\gamma_n}}}}}. \tag{4}$$

Here the s_i are integers, and γ_n is an irrational number > 1 chosen so that equality holds. Then s_n is the greatest integer $\leqslant \gamma_n$. *If in* (4) γ_n *is replaced by* s_n, *the fraction chain becomes a rational number* p_n/q_n. *These* p_n/q_n *are the terms in the (fast) continued fraction algorithm for* α; *they are left approximations* a/b *if* n *is even, and right approximations* c/d *otherwise*. The variable n, which we suppressed in our discussion of the fast continued fraction algorithm, indexes the rightward and leftward migrations in our geometric presentation.

To prove these statements, we will use the computer algorithm as our pivot. We have already seen that the geometric theory leads to this algorithm; now we deduce the same algorithm from the fraction chain (4), thus proving the equivalence of the two theories.

First, it is easy to prove by induction that, given α, a unique fraction chain (4) exists for every n. For, assuming this holds for n, we immediately deduce the recursion

$$\gamma_{n+1} = 1/(\gamma_n - s_n).$$

We notice that this matches formula (3) in the computer algorithm. Now our objective is to recover formulas (1) and (2). This requires another induction.

We think of the fraction chain (4) as a function of γ_n, with the s_i held fixed and γ_n varying. In order to prove the assertion in italics, we will show that the function (4) has the form:

$$\frac{p_{n-2} + \gamma_n p_{n-1}}{q_{n-2} + \gamma_n q_{n-1}}. \tag{5}$$

(To match the notation of the computer algorithm we would write $a/b = p_{n-2}/q_{n-2}$ and $c/d = p_{n-1}/q_{n-1}$.) Now to make the induction from n to $n+1$, we recall that p_n/q_n is the result of replacing γ_n by s_n in (4), and that $\gamma_{n+1} = 1/(\gamma_n - s_n)$. Thus:

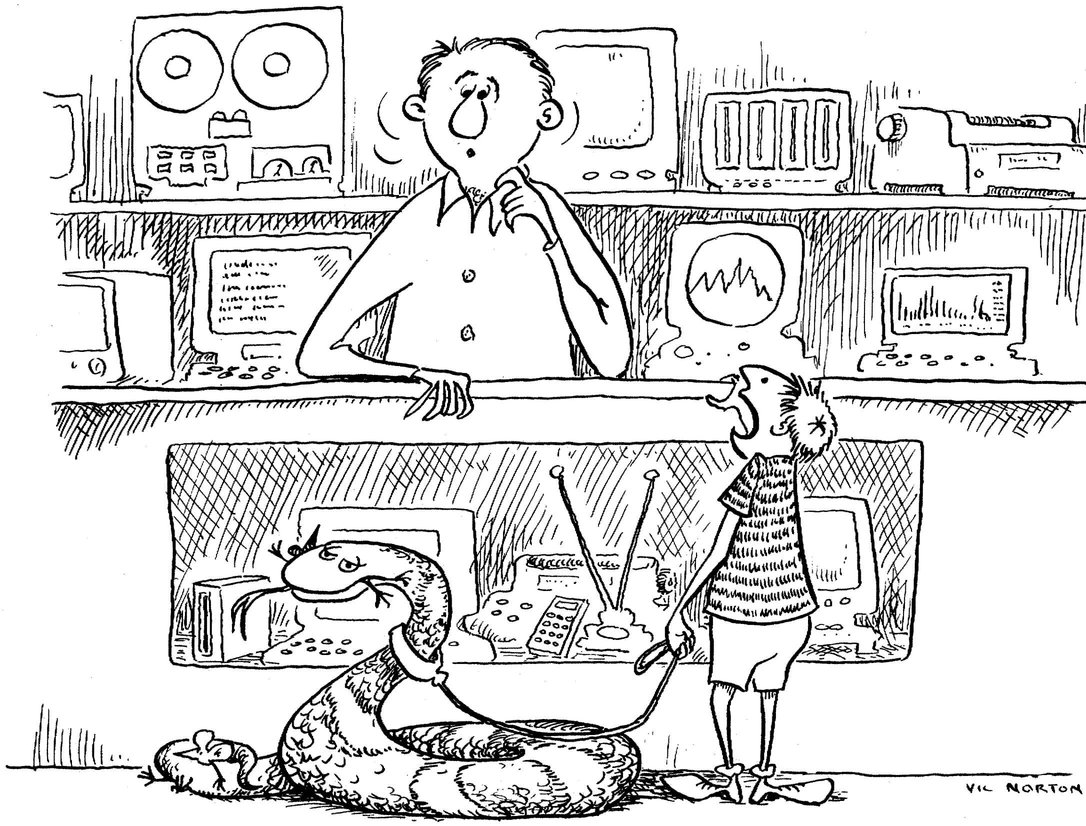
$$\frac{p_{n-1} + \gamma_{n+1} p_n}{q_{n-1} + \gamma_{n+1} q_n} = \frac{p_{n-1} + \gamma_{n+1}(p_{n-2} + s_n p_{n-1})}{q_{n-1} + \gamma_{n+1}(q_{n-2} + s_n q_{n-1})},$$

and using $1/\gamma_{n+1} = \gamma_n - s_n$ yields

$$\frac{(\gamma_n - s_n)p_{n-1} + (p_{n-2} + s_n p_{n-1})}{(\gamma_n - s_n)q_{n-1} + (q_{n-2} + s_n q_{n-1})} = \frac{p_{n-2} + \gamma_n p_{n-1}}{q_{n-2} + \gamma_n q_{n-1}}.$$

This completes the induction and establishes formulas (1) and (2).

Thus our previous theory, in its geometric setting, is equivalent to that determined by the endless fraction.



I WANT TO TRADE MY PYTHON FOR AN ELECTRONIC **FAREY** PING-PONG GAME.

Suggestions for further reading

An interesting geometrical approach to continued fractions is given in the book by Harold Stark [**4**]. Roughly speaking, Stark's approach is two dimensional (based on the slopes of lines), whereas our approach is one dimensional. For the classical "infinite fraction chain" viewpoint, see almost any book on elementary number theory. My favorite is Hardy and Wright [**2**]. An elementary introduction to fractions (Egyptian fractions, Farey fractions, continued fractions, decimal fractions) can be found in [**1**].

References

[1] A. Beck, M. Bleicher, D. Crowe, Excursions into Mathematics, Worth, NY, 1969, pp. 414–475.
[2] G. H. Hardy and E. M. Wright, The Theory of Numbers, 4th ed., Oxford Univ. Press, 1960.
[3] M. Shrader-Frechette, Modified Farey sequences and continued fractions, this MAGAZINE, 54 (1981) 60–63.
[4] H. Stark, An Introduction to Number Theory, Markham, Chicago, 1970.

Boole's Algebra Isn't Boolean Algebra

A description, using modern algebra, of what Boole really did create.

THEODORE HAILPERIN
Lehigh University
Bethlehem, PA 18015

To Boole and his mid-nineteenth century contemporaries, the title of this article would have been very puzzling. For Boole's first work in logic, *The Mathematical Analysis of Logic*, appeared in 1847 and, although the beginnings of modern abstract algebra can be traced back to the early part of the nineteenth century, the subject had not fully emerged until towards the end of the century. Only then could one clearly distinguish and compare algebras. (We use the term **algebra** here as standing for a formal system, not a structure which realizes, or is a model for, it—for instance, the algebra of integral domains as codified by a set of axioms *versus* a particular structure, e.g., the integers, which satisfies these axioms.) Granted, however, that this later full degree of understanding has been attained, and that one can conceptually distinguish algebras, is it not true that Boole's "algebra of logic" is Boolean algebra?

Conventional opinion without exception is on the affirmative side of this question. Briefly put, this opinion has it that Boole was the first to give a successful algebraic setting for doing logic (of class terms, or properties) and, as is well known, the abstract form of the calculus of classes (classes = sets), is Boolean algebra. But if we look carefully at what Boole actually did—his major work is *The Laws of Thought* of 1854—we find him carrying out operations, procedures, and processes of an algebraic character, often inadequately justified by present-day standards and, at times, making no sense to a modern mathematician. Not only that, but even though his starting algebraic equations (expressing premises of an argument) were meaningful when interpreted in logic, the allowed transformations often led to equations without meaning in logic. Boole considered this acceptable so long as the end result could be given a meaning, which he did give by a specifically introduced procedure. Writing in 1870 the logician (and economist) W. S. Jevons complained:

> ...he (Boole) shrouded the simplest logical processes in the mysterious operations of a mathematical calculus. The intricate trains of symbolic transformations, by which many of the examples in the Laws of Thought are solved, can be followed only by highly accomplished mathematical minds; and even a mathematician would fail to find any demonstrative force in a calculus which fearlessly employs unmeaning and incomprehensible symbols, and attributes a signification to them by a subsequent process of interpretation [**4**, p. 143].

As this quote indicates, it appears that to determine what sort of algebra Boole did use to do logic will require a fairly substantial exegetical effort, and what the outcome will be, whether Boolean algebra or something else, is not immediately clear. Before entering into a presentation of Boole's ideas, we depict the foundations-of-algebra *milieu* in which he worked.

Symbolical Algebra (circa 1830-1840)

Although in 1837 W. R. Hamilton had shown how to define complex numbers as ordered pairs of ordinary numbers, there was, as yet, no real understanding of the various number systems as being successive, more inclusive, extensions of natural numbers; nor was there any conception of such number structures as models of appropriate formal algebraic systems. Such fundamental algebraic properties as commutativity and distributivity had only recently been so named (by F-J.

GEORGE BOOLE, F.R.S.

Engraving reproduced from "George Boole, F. R. S. 1815–1864," by G. Taylor, *Notes and Records of the Royal Society of London,* (12) 1956, 44-52.

Servois in 1814/15), and associativity not until 1844 (by Hamilton). And, despite the fact that over the course of several centuries the techniques and application of numerical algebra had become highly developed and refined, there was no adequate body of justifying principles. Symbolical Algebra, as it was called, claimed to provide these absent principles. Its chief promulgator was George Peacock, Fellow and Tutor of Trinity College, Cambridge. Also prominent in its advocacy were Duncan F. Gregory, editor of the journal which published Boole's earliest mathematical research, and Augustus De Morgan, an early champion of Boole's work in logic, who also wrote extensively on formal logic.

The doctrine espoused by Peacock was that there were two sciences of algebra, *arithmetical* and *symbolical*. In the former the general symbols and operations refer to the numbers and operations of "common arithmetic" whose meaning required in many cases restriction on the performability of the operations. For example, in arithmetical algebra one couldn't subtract a larger from a smaller number and hence the form $a-b$ involving the general symbols a and b could be meaningless if so interpreted. In symbolical algebra, however, all restrictions on operations are removed—what meaning expressions then had was to be subsequently determined from the assumed laws of symbols. The laws are obtained from the following principles (stated in Peacock's language to retain historicity):

(i) Whatever forms in general symbols are equivalent in arithmetical algebra, are also equivalent in symbolical algebra.

(ii) Whatever forms are equivalent in arithmetical algebra when the symbols are general in their form, though specific in their value, will continue to be equivalent when the symbols are general in their nature as well as in their form.

Thus although $a(b-c)=ab-ac$ is true in arithmetical algebra only if b is not less than c, this restriction is removed in symbolical algebra and the equation is considered to be true. Symbolical algebra is to be an extension of the arithmetical in the sense that when its symbols are numbers and its operations arithmetical, the results must be identical with that of arithmetic. However, "...inasmuch as in many cases, the operations required to be performed are impossible, and their results inexplicable, in their ordinary sense, it follows that the meaning of the operations performed, as well as the results obtained under such circumstance, must be derived from the assumed rules, and not from their definitions or assumed meanings, as in Arithmetical Algebra." [**9**, p. 7] In contrast to this view, contemporary mathematics does not consider formal rules (or axioms) as being able to create meanings for operations, only to set limits for them—e.g., adoption of $a+b=b+a$ excludes the symbol $+$ from being interpreted as a noncommutative operation. It should be emphasized that nowhere in Peacock's *Treatise* do we find a full listing of algebraic properties of the operations analogous to a present day set of axioms.

Despite such inadequacy, and much more that one could find fault with, symbolical algebra did provide an environment in which algebra could be freed from its exclusive use with number. For example, it provided a measure of justification for the calculus of operations or separation of symbols, as it was also known. In this calculus one separates off the symbols of operation and performs algebraic operations on them. A simple example is the well-known symbolic formula $(D+a)^2=D^2+2aD+a^2$ of elementary differential equations in which D stands for the operation of differentiation and the square on D indicates a two-fold application. This type of procedure, of applying algebra to "symbols" and using special interpretation, was suspect in the view of some mathematicians—Cauchy for one—but was widely used in England. Boole was a leading practitioner of the art.

Though a confirmed adherent of the symbolical algebra school of thought, Boole was no rigid follower. He significantly modified Peacock's principles by relaxing the requirement that when the general symbols and the operations have their common arithmetical meaning, the result should be an arithmetical truth. For example, in a paper of 1844 for which he received a Royal Society medal, Boole introduced noncommuting types of operators. Turning to the immediate subject at hand, the characteristic law $x^2=x$ which he added to algebra so as to do logic is not true arithmetically save—as Boole points out—if one restricts x to being 0 or 1. Did Boole thus create

Boolean algebra by adding $x^2=x$? Before addressing ourselves to the question we turn to a brief exposition of Boolean algebra.

Boolean Algebra

Nowadays we describe a formal algebraic system by way of axioms. For Boolean algebra especially, such axiom systems come in a large variety of shapes (basic operations, relations) and sizes (number of axioms). The one we choose to present here focuses attention on analogies with numerical algebra. It uses two binary operation symbols, $+$ and $\times$, one unary operation symbol $'$, and two constant symbols 0 and 1.

Axioms for Boolean Algebra

(i)	$ab=ba$,	$a+b=b+a$,
(ii)	$(ab)c=a(bc)$,	$(a+b)+c=a+(b+c)$,
(iii)	$1a=a$,	$0+a=a$,
(iv)	$aa'=0$,	$a+a'=1$,
(v)	$a(b+c)=ab+ac$,	$a+(bc)=(a+b)(a+c)$,
(vi)	$aa=a$,	$a+a=a$.

The evident analogy with ordinary algebra which is displayed here can be further enhanced by the introduction of a subtraction symbol in terms of which $'$ can be expressed: we let $x-y$ stand for xy' so that, with (iii), we have $1-x=1x'=x'$. Hence in place of (iv) we can write

$$\text{(iv*)}\quad a(1-a)=0, \qquad a+(1-a)=1.$$

We need to distinguish **a Boolean algebra** from the *general concept of*, or the *formal theory of*, Boolean algebras. When using the indefinite article, we are referring to a particular mathematical structure which, via an appropriate interpretation, satisfies the above axioms. The best known example of a Boolean algebra is the set of subsets of a fixed set, with the set operations of union, intersection and complementation interpreting, respectively, the symbols $+$, $\times$, $'$, and with the empty set and the fixed set interpreting 0 and 1.

One can come to Boolean algebra from another direction which adds a little more to the analogy with numerical algebra. The Boolean addition in the above axioms lacks the property of linear solvability or, equivalently, existence of an additive inverse; that is, $a+x=b$ need not have a solution for x nor, if it did, a unique one. If, however, one introduces by definition an operation $+_\Delta$ ("symmetric difference") by putting

$$x+_\Delta y=x-y+y-x=xy'+yx',$$

then it is true that in any Boolean algebra the equation $a+_\Delta x=b$ has a unique solution for x, for any a and b. One can show that, under the operations $+_\Delta$ and $\times$, with 0 and 1 as the zero and unit, a Boolean algebra is a commutative ring with unit. A commutative ring having the idempotency property $a^2=a$ is called a **Boolean ring**. It is straightforward to show that the theories of Boolean algebra and Boolean rings (with unit) are equivalent. The particular axiom system given above for Boolean algebras has the property that it becomes an axiom system for Boolean rings just by replacing the axiom $a+a=a$ by $a+a=0$ (with all the $+$'s then taken to be symmetric difference). A well-known simple example of a mathematical structure which is a Boolean ring is the field of integers modulo 2.

The close affinity of any Boolean algebra with two-valued arithmetic is expressed, in sophisticated language, by the following representation theorem [**10**, p. 50]: *every (non-degenerate) Boolean algebra is isomorphic to a subalgebra of a direct union of two-element Boolean algebras*. In unsophisticated language this says that any Boolean algebra "looks like" an algebra (Boolean, of course) on a subset of all n-tuples $U=\{\langle v_1,v_2,\cdots\rangle\}$ with each v_i either 0_i or 1_i (the zero and unit of a two-element Boolean algebra), and with operations on the n-tuples defined componentwise. The importance of the two-valued arithmetic in connection with Boolean algebras is also brought out in the following useful result: any Boolean polynomial equation in n variables is true for all Boolean algebras (i.e., is a Boolean identity) if true for all 2^n possible assignments of 0's and 1's for the variables. We also note that by virtue of idempotency ($x^2 = x$) a Boolean polynomial is linear in each of its variables.

Since a Boolean algebra can be viewed as a ring (with symmetric difference as the addition) the notions of ideals and of residue classes can be introduced in customary fashion.

Boole's Logic of Class Terms

Boole believed he was the first to use a mathematical approach to logic. But much earlier Leibniz had conceived of the idea of a formal mathematical system which could be used to conduct logical inference. In manuscript notes and papers—which did not become generally known until some 200 years later—he had made some remarkable starts but did not bring the project to fruition. One of the minor hampering items may have been his regarding general terms (e.g., "human," "sheep") primarily as designating attributes rather than classes (i.e., thinking *intensionally* rather than *extensionally*). In this view the compound term "human sheep" has a wider, more inclusive, intension than either "human" or "sheep," whereas as a class it has a less inclusive extension than either of the component terms. Thinking extensionally is far simpler; for example, the terms "human" and "featherless biped" designate distinct attributes but as classes are identical (plucked chickens being ruled out). Leibniz could, and did, think either way, but Boole's thinking was exclusively extensional.

A more substantial difference, however, was the circumstance that Leibniz, faced with *infiniti modi calculandi*, had to grope his way towards a calculus of logic, whereas Boole was sure he had one ready-made, namely symbolical algebra, furnished with a suitable interpretation for its general symbols and operations. Having as a guide a theoretical system, even if only approximately correct, can be decidedly advantageous. As we now know, neither of them had at their disposal a clear and sufficiently extensive body of formal logical usage—what was available, to Boole as well as to Leibniz, wasn't much more than the Aristotelian syllogistic, a body of rules not lending itself well to formulation as an algebra of the equational type, and, moreover, burdened with the distracting problem of existential import (Does "All S are P" imply "There are S?").

In his initial writings on logic Boole associated the general symbols $x, y, z, \ldots$ of symbolical algebra with *operators* which selected classes out of a universal class. (Subsequently, this aspect became less pronounced and the symbols came to stand for the classes themselves.) Selecting out the x's and then the y's produces the same class as selecting out the y's and then the x's. Thus Boole wrote

$$xy = yx.$$

For Boole to think of using multiplication for logical intersection was quite natural since this was what was used in the calculus of operators to indicate successive application of operations. In keeping with this "selecting out" idea the symbolic laws

$$x \cdot 1 = x, \quad x \cdot 0 = 0$$

inevitably suggest the logical interpretation for 1 as the universal class and 0 as the empty class. Selecting out the x's and then from this the x's yields the x's; hence the algebraic property,

$$xx = x^2 = x,$$

a departure from ordinary algebra which we have already mentioned.

To symbolize the aggregate of two classes (e.g., "mountains and vales") Boole used the symbol of addition but with the proviso that the classes be disjoint and hence the operation is only partially defined. Boole had no objection to partially defined operations—compare, for instance, $a - b$ in arithmetical algebra. Similarly, subtraction required that the "subtrahend" class be contained in the "diminunend" class. Under these conditions he found

$$x(y+z) = xy + xz$$

and

$$x(y-z) = xy - xz$$

are true for classes. With hardly much more than these examples Boole boldly proposed [**1**, p. 37]:

> Let us conceive, then, of an Algebra in which the symbols, x, y, z, etc., admit indifferently of the values 0 and 1, and of these values alone. The laws, the axioms, and the processes, of such an Algebra will be identical in their whole extent with the laws, the axioms, and the processes of an Algebra of Logic. Difference of interpretation will alone divide them. Upon this principle the method of the following work is established.

In these sentences we have, for the period, some remarkably innovative ideas: that there can be an algebra of 0 and 1, that such an algebra has at least two interpretations, and that one of these is logic. These items give rise to the following questions.

(1) What is Boole's algebra of 0 and 1?
(2) Is arithmetic modulo 2 one of the interpretations?
(3) How does one interpret the algebra to get logic?

It will be convenient for us to address these questions in reverse order.

Re (3). As already indicated, the symbols x, y, z, etc., are to stand for classes, with 1 being the universe and 0 the empty class. The product xy represents the intersection of the classes represented by x and y and is meaningful in all circumstances. The aggregate class $x+y$ has meaning only if x and y are mutually exclusive; likewise $x-y$ only if y is contained in x (and thus is *not* the same as the xy' of Boolean algebra). The expression $1-x$ is meaningful in any case and represents the complement of x in the universe 1. Assertions are always represented by equations—e.g., "No x's are y's" is "$xy=0$" and "All x's are y's" is "$x=vy$" where v is an indefinite class symbol. This use of the special symbol v was adversely criticized by other logicians but can in part be justified by considering it as an understood existentially quantified variable [**2**, pp. 97–98]. Logical inference—about which we will say very little in this paper—is performed by algebraic transformations on equations, allowing all the operations of "common" algebra. This creates a problem (to be discussed in detail in our next section) in that expressions arise, e.g., $-1, x+x, x-1, x/y$, to which no logical meaning has been assigned. Boole considered this perfectly acceptable in intermediate steps of a deduction so long as the starting and ending equations were interpretable in logic. He likened this to "the employment of the uninterpretable symbol $\sqrt{-1}$ in the intermediate processes of trigonometry." Aside from this dubious feature, the direct translating of logical statements involving class terms *into* algebraic form goes tolerably well.

Re (2). It is well known that an interpretation for Boolean algebra (i.e., the algebra of section 2) can be given as an algebra over the two-element set $\{0,1\}$ whose arithmetic differs from ordinary arithmetic in that one has either (i) $1+1=1$ or (ii) $1+1=0$, the latter case yielding arithmetic modulo 2. However, neither (i) nor (ii) can be the interpretation Boole has in mind, for he always writes "$1+1$" as "2", considering 2 as a "numerical" factor which does not obey the general law of thought $x^2=x$. Correspondingly, for any algebraic expression A, the sum "$A+A$" is written "$2A$" and never simplified. In a letter to Jevons written in 1863 in response to an inquiry, Boole flatly asserted that $x+x=x$ was not true (i.e., a law) in logic, so that (by implication) $1+1=1$ would not be true in his algebra of 0 and 1. Clearly then Boole would consider neither arithmetic modulo 2 nor the alternative with $1+1=1$ as an interpretation for his algebra.

Re (1). It is not easy to determine what Boole's algebra is since he gives no explicit list of its laws (i.e., axioms). He talks about $x^2=x$ being the general law of thought, but $A^2=A$ for arbitrary A is not one of his algebraic laws—it holds for *class symbols* x, y, etc., and for expressions such as

$$xy, 1-x, xy'+x'y$$

which he refers to as "independently interpretable," but not for expressions such as

$$2x, x+y, x-y$$

which nevertheless do occur in his procedures. An expression A not satisfying the idempotency condition $A^2=A$ is dubbed "uninterpretable." A careful examination of those algebraic properties of $+$ and $\times$ which Boole actually makes use of, for interpretables as well as uninterpretables, show that they are those of a

commutative ring with unit having no additive or multiplicative nilpotents. **(SM)**

(No additive or multiplicative nilpotents means that $nA = 0 \Rightarrow A = 0$ and that $A^n = 0 \Rightarrow A = 0$.) To use this algebra for logic (of class terms) one needs entities which satisfy Boole's law of thought $x^2 = x$. Let us single out and call such entities the Boolean (or idempotent) elements and designate the set of such elements by **B**. Clearly the 0 and 1 of the ring are in **B**. Also, for any $x, y \in \mathbf{B}$, the commutative ring properties imply that the elements

$$xy, 1 - x, xy' + x'y, \quad \text{and} \quad xy + x'y + xy'$$

(here $x' = 1 - x$) are also in **B**. It is easy to show that the elements of **B** will constitute a Boolean algebra (or equivalently a Boolean ring) with the operation

$$x \cup y = xy + x'y + xy'$$

as the Boolean sum, or with

$$x +_{\Delta} y = x'y + xy'$$

as the Boolean ring addition. Thus by restricting himself to the idempotents of his algebra, Boole would have had Boolean algebra—and we rightfully honor him by attaching his name to this algebra. Only *Boole* didn't know it. He steadfastly refused to acknowledge any operation but his $+$. When Jevons claimed that $x + x = x$, and Boole emphatically denied this, they were really talking about different operations. Jevons' $+$ is indeed the present day $\cup$; however Boole's $+$ is not, as is generally believed, the $+_{\Delta}$ of the Boolean ring but the $+$ of the ring **SM**. This algebra merits a brief discussion.

What kinds of models (realizations) does Boole's algebra **SM** have? An answer to this question comes from a structure theorem of N. H. McCoy [**8**, p. 123] which, adapted to our circumstance reads: any model of **SM** is isomorphic to a subdirect sum of integral domains which are without additive nilpotents. Intuitively such a model looks like n-tuples of elements in which each component ranges over an integral domain without nilpotents, and with the operations of $+$ and $\times$ for these n-tuples defined componentwise.

Not all models or interpretations of a set of axioms are of equal general interest. In the case of Boolean algebra there is a sense in which one can say that the principal interpretation is that of an algebra of sets. For Boole's algebra **SM** we consider the principal interpretation to be that of an **algebra of signed multisets**: a multiset is like an ordinary set except that multiple occurrences of elements are allowed, and by a signed multiset we mean one in which negative multiplicities are allowed. For example, during a poker game your pile of chips contains various positive or zero multiples of red, white or blue chips, all of each color indistinguishable as far as the game is concerned; and if you borrow from the pot, you are adding negative multiples to your pile. When the multiplicities are restricted to being either 0 or 1, then multisets become ordinary sets. As for the operations, Boole's $+$ corresponds to dumping the contents of the two multisets together, and $\times$ corresponds to multiplying the respective multiplicities. (For more details see [**2**, pp. 91, 92–96]; in that work we used the term "heap," not realizing that "multiset" was already in use.)

If all one wanted to use Boole's algebra for was to do class (i.e., term) logic, then there is no need to go beyond the idempotents—and this is the path history chose. Nevertheless, the richer structure of Boole's algebra can be of interest in its own right, for example, in the logic of multisets. As a simple instance, the equivalence

$$(x + x)y = 0 \Leftrightarrow xy = 0$$

when interpreted for multisets tells us that duplicating the elements of a set does not affect its being exclusive from another set. It is possible that Boole's algebra can be useful in the subject of pseudo-Boolean functions [**3**]. Pseudo-Boolean functions, a notion used in operations research, are functions from a two-element set $\{0, 1\}$ into the integers, and they have an obvious representation by means of a polynomial in Boolean variables with integer coefficients.

Boole's Interpretation Procedure

As we have mentioned earlier, Boole considered his algebraic methods for doing logic to be sound so long as the starting and end formulas were interpretable in logic. But many of the algebraic processes—especially division, which plays an important role in Boole's method—led to uninterpretable forms. For example, in order to determine what the Boolean equation $xw=y$ implies about the class w, Boole solves this equation for w, obtaining

$$w=\frac{y}{x}.$$

What does y/x mean in logic? What is the interpretation for this equation? Boole ignores the first of these questions, but to handle the second he introduces the idea of the **expansion** (or **development) of a function** and along with it a general method for interpreting an equation between a class symbol and such an expansion. By this technique—widely decried as "mysterious"—Boole had a general and uniform method for obtaining the logical import of any equation in class symbols (i.e., Boolean variables). We give a brief description of this procedure.

If $f(x)$ is an algebraic expression (Boole only considered linear fractional forms in Boolean variables), its expansion is given by

$$f(x)=f(1)x+f(0)(1-x).$$

Boole "establishes" this identity by assuming $f(x)=ax+b(1-x)$ and determining a and b by setting $x=1$ and $x=0$. Similarly (using x' as an abbreviation for $1-x$) for an expression in two variables he gives

$$f(x,y)=f(1,1)xy+f(1,0)xy'+f(0,1)x'y+f(0,0)x'y'.$$

Thus if $f(x,y)=\frac{y}{x}$, then

$$\frac{y}{x}=1\cdot xy+0\cdot xy'+\frac{0}{0}x'y'+\frac{1}{0}x'y.$$

(We follow Boole's custom of putting the $1/0$ term last.) Boole's argument for equating a function with its expansion is faulty in that it assumes (without justification) that any such function is linear in its Boolean variables. To continue, Boole's rule for interpreting an equation $w=f(x,y,\ldots)$ is to equate w to the sum of those terms in the expansion of $f(x,y,\ldots)$ which have 1 as their coefficient plus an indefinite multiple of the sum of those terms having $0/0$ as their coefficient and, as an independent condition, the sum of those terms having $1/0$ as coefficient is equated to 0. The reasons Boole gives for his rule are weak and unconvincing and we shall not reproduce them. As an example of its use, the interpretation for the equation $w=y/x$ is that the class w consists of all of xy, none of xy', an indefinite amount (some, none, or all) of $x'y'$ and, independently, $x'y$ is set equal to 0. The correctness of this as an equivalent to $xw=y$ is easily seen in FIGURE 1.

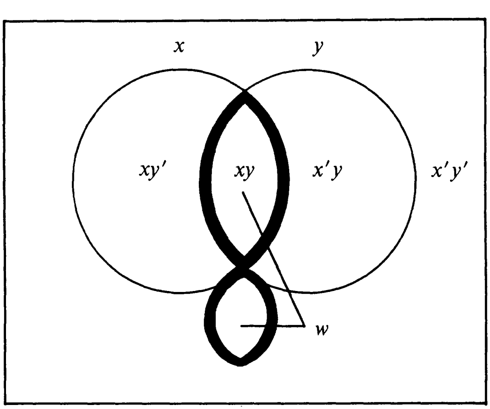


FIGURE 1.

$xw=y$

equivalent to

$$\begin{cases} x'y=0 \\ w=xy+vx'y',\ v\ \text{indefinite.}\end{cases}$$

In the *Laws of Thought* there are lots of worked-out examples, all with correct conclusions. One is intrigued by this fact and wonders how such a process involving "division" of class expressions, and using entities such as $0/0$ and $1/0$, can be meaningful. We venture an explanation for this in our next section.

Division in Boolean Algebras

In the arithmetic of integers the quotient of m by n exists if $n \neq 0$ and m is a multiple of n. Thus for integers the operation of division is a restricted one and is not always possible. The process of extending the integers to the more extensive structure of the field of rationals is well known: one defines the rationals as equivalence classes of ordered pairs of integers (here ordered pairs = fractions) and with appropriate definitions of addition, multiplication, zero and unit, one has a new structure in which division is possible without exception (save for 0). This structure, moreover, contains a substructure isomorphic to the integers, so that one can still do integer arithmetic within the rationals.

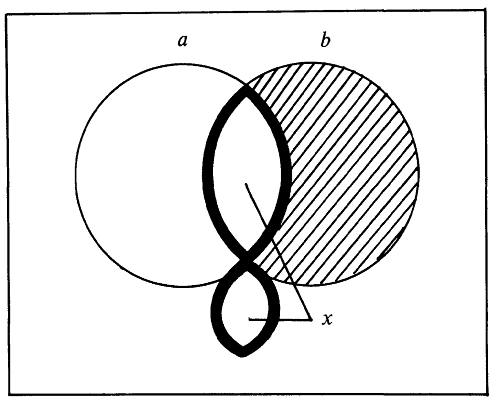


FIGURE 2

Consider now in a Boolean algebra $\mathbf{B}$ the equation

$$ax = b, \tag{1}$$

where a and b are given and x is unknown. It is a simple exercise in Boolean algebra to show that this equation has a solution for x if and only if b is included in a, i.e., if and only if $a'b = 0$. Thus, as with the integers, quotients do exist for some pairs of elements of $\mathbf{B}$. However, unlike the integers, such a quotient need not be unique. As we have already seen in illustrating Boole's interpretive procedure, a solution x for $ax = b$ will be equal to ab plus any amount of $a'b'$ (see FIGURE 2); or, to say it in other words, any two members of the solution set for (1) will differ by a Boolean multiple of (i.e., a part of) $a'b'$. We now show how to "factor out" the differences so as to have a unique quotient (if any exist).

To say that any two solutions of (1) differ by a multiple of $a'b'$ is to say that the solution set is a residue class in the factor ring $\mathbf{B}/(a'b')$, where $(a'b')$ is the principal ideal generated by $a'b'$ (= set of classes included in $a'b'$). If in the factor ring $\mathbf{B}/(a'b')$ we denote by

$$[r] = r + (a'b')$$

the residue class determined by an element r of $\mathbf{B}$, then the mapping $r \to [r]$ is a homomorphism of rings and hence (1) implies that

$$[a][x] = [b] \tag{2}$$

holds in the ring $\mathbf{B}/(a'b')$. Conversely, (2) implies (1). For if, for $v_1, v_2, v_3 \in \mathbf{B}$, we have

$$(a+v_1a'b')(x+v_2a'b')=b+v_3a'b',$$

then by multiplying this equation through by $(a'b')'(=a\cup b)$ and using simple Boolean identities, one readily obtains (1). Thus the problem of solving for x in (1) is equivalent to finding $[x]$ satisfying (2) in the reduced Boolean algebra $\mathbf{B}/(a'b')$. If there is a solution, say $[x]=[p]$, then, on using the fuller notation for a residue class, this solution is

$$[x]=p+(a'b'). \tag{3}$$

In order to better bring out the relation between what we are doing with Boole's procedure, we now alter the customary notation for a principal ideal in a ring and write $\frac{0}{0}a'b'$ in place of $(a'b')$ so that (3) becomes

$$[x]=p+\frac{0}{0}a'b' \tag{4}$$

which, in terms of elements of $\mathbf{B}$, says

$$x=p+va'b' \tag{5}$$

where v is an element of $\mathbf{B}$. By going over to $\mathbf{B}/\frac{0}{0}a'b'$ we are now at a stage similar to solving an equation $mx=n$ with m and n integers, where only under limited circumstances does a (unique) quotient exist. We would like the greater algorithmic freedom that one has when working with rationals where division is unrestricted. This is essential if we are to reach our goal of justifying Boole's unusual technique, which involves solving equations such as

$$Ew=F$$

for the unknown w with E and F polynomials in variables $x,y,z,\ldots$. To this end we turn to the problem of extending a Boolean algebra by the introduction of quotients.

Extending a Boolean Algebra with Quotients

In contemplating the introduction of quotients and the operation of division into a Boolean algebra, one is confronted with the difficulty that the usual definitions of addition and multiplication of two fractions, namely

$$\frac{r_1}{s_1}+\frac{r_2}{s_2}=\frac{r_1s_2+s_1r_2}{s_1s_2}$$
$$\frac{r_1}{s_1}\cdot\frac{r_2}{s_2}=\frac{r_1r_2}{s_1s_2},$$

involves the product of the two denominators. In the case of the integral domain of integers, where there are no divisors of zero, if s_1 and s_2 are not 0, then neither is the product s_1s_2. However in the case of Boolean algebra, where $aa'=0$ for any a, *every* nonzero element is a divisor of zero. Thus no element of a Boolean algebra can function with complete freedom as a denominator.

Resolution of the difficulty (though not without a price) comes by suitably restricting the denominators to being in a multiplicative set: we say S is a **multiplicative set** if

(i) $0\notin S$ and
(ii) $s_1,s_2\in S\Rightarrow s_1s_2\in S$ for any s_1, s_2.

For any commutative ring R with unit, by restricting the formation of fractions to those having denominators coming from a multiplicative subset S of R, one can alleviate the aforementioned difficulty. But now the customary definition of the equivalence of two fractions $\left(\frac{r_1}{s_1}\sim\frac{r_2}{s_2}\ \textit{iff}\ r_1s_2-s_1r_2=0\right)$ has to be modified to

$$\frac{r_1}{s_1}\sim\frac{r_2}{s_2}\text{ iff for some } s\in S,\ s(r_1s_2-s_1r_2)=0. \tag{6}$$

Denote by $\left[\frac{r}{s}\right]$ the equivalence class determined by a pair of elements $r, s \in R$, $s \in S$, and by RS^{-1} the set of all such equivalence classes. Then with operators $+$ and $\times$ defined by

$$\left[\frac{r_1}{s_1}\right]+\left[\frac{r_2}{s_2}\right] = \left[\frac{r_1 s_2 + s_1 r_2}{s_1 s_2}\right],$$
$$\left[\frac{r_1}{s_1}\right]\left[\frac{r_2}{s_2}\right] = \left[\frac{r_1 r_2}{s_1 s_2}\right],$$

and with $\left[\frac{0}{1}\right]$ and $\left[\frac{1}{1}\right]$ as the zero and unit, one can show that RS^{-1} is a ring—the ring of quotients of R by S [**12**]. And now here is the price we have to pay in order to have quotients in a general ring: in the resulting ring of quotients not every member of R can be a denominator, nor need the ring of quotients be a field; moreover, although the mapping $h: R \to RS^{-1}$, $h(r) = \left[\frac{r}{1}\right]$, is a homomorphism of rings, it need not be injective—as would be the case for the rationals.

To apply this theory to a Boolean algebra (which we take in the form of a ring with $+$ standing for symmetric difference) one chooses as a denominator set any filter. If e is a nonzero element of a Boolean algebra $\mathbf{B}$, then the filter determined by e consists of all elements which include e. We may conveniently designate this set by $e \cup \mathbf{B}$. Clearly $e \cup \mathbf{B}$ is a multiplicative set and thus $\mathbf{B}(e \cup \mathbf{B})^{-1}$ is a ring of quotients of $\mathbf{B}$ by $e \cup \mathbf{B}$. As the usual notation for quotients is here insufficient in that it contains no indication of the denominator set, let us use $\left(\frac{r}{s}\right)_e$ where the e suffices to indicate the denominator set $e \cup \mathbf{B}$.

There are a few results about Boolean quotients to which we will refer:

THEOREM 1. *For any element a in the denominator set $e \cup \mathbf{B}$*

$$\text{(i)} \left(\frac{ab}{a}\right)_e = \left(\frac{b}{1}\right)_e,$$
$$\text{(ii)} \left(\frac{b}{a}\right)_e = \left(\frac{ab}{1}\right)_e.$$

Proof. Both results are immediate consequences of definition (6); (i) by virtue of $e(1 \cdot ab - a \cdot b) = 0$, and (ii) by $a(b \cdot 1 - a \cdot ab) = 0$.

THEOREM 2. *The factor ring $\mathbf{B}/(e')$ and the ring of quotients $\mathbf{B}(e \cup \mathbf{B})^{-1}$ are, under the mapping $b + (e') \to \left(\frac{b}{1}\right)_e$, isomorphic structures.*

Proof. [**2**, p. 34].

We now show how to reproduce, with due mathematical rigor, Boole's solution of a Boolean equation by division and expansion, and to justify his interpretation for the algebraic solution. Since the solution of a Boolean equation is directly obtainable in Boolean algebra, our procedure, which uses a number of new structures, will appear complicated and circuitous. We emphasize, however, that it is the reproducing of Boole's method, not the solution of the equation, which is of interest.

Consider an equation

$$ax = b \tag{7}$$

where $a (\neq 0)$ and b are elements of a Boolean algebra $\mathbf{B}$. As we have seen above, this equation has a solution for x in $\mathbf{B}$ if and only if the equation

$$[a][x] = [b] \tag{8}$$

has a solution for $[x]$ in $\mathbf{B}^* = \mathbf{B}/\frac{0}{0}a'b'$, where $\frac{0}{0}a'b'$ is the principal ideal generated by $a'b'$ and $[r]$ designates the residue class modulo $\frac{0}{0}a'b'$ determined by r; moreover, if $[x] = [p]$ is the solution of (8), then the solution set for (7) is given by

$$x = p + va'b', \quad v \text{ ranging over all elements of } \mathbf{B}.$$

Suppose now that (8) has a solution. In the ring of quotients $\mathbf{Q} = \mathbf{B}^*([a] \cup \mathbf{B}^*)^{-1}$ we have, from (8),

$$\left(\frac{[a][x]}{[a]}\right)_a = \left(\frac{[b]}{[a]}\right)_a, \tag{9}$$

from which, by cancellation (Theorem 1(i)),

$$\left(\frac{[x]}{[1]}\right)_a = \left(\frac{[b]}{[a]}\right)_a. \tag{10}$$

This is what for us corresponds to Boole's $x = b/a$. Using Theorem 1(ii) and the fact that the mapping $r \to [r]$ is a homomorphism to replace $[a][b]$ by $[ab]$, we obtain from (10) that

$$\left(\frac{[x]}{[1]}\right)_a = \left(\frac{[ab]}{[1]}\right)_a \tag{11}$$

in the ring $\mathbf{Q}$. We now go over to the ring $\mathbf{B}^*/([a]')$ which, by Theorem 2, is isomorphic to $\mathbf{Q}$. But first we note that $[a]' = [a']$, and that $[a'] = [a'b]$ since a' and $a'b$ differ by $a'b'$ and thus determine the same residue class modulo $a'b'$. Hence the ideal $([a]')$ is $([a'b])$. We change the usual notation and represent this ideal by $\frac{1}{0}a'b$. (Note that $\frac{1}{0}a'b$ is an ideal of $\mathbf{B}^*$, whereas $\frac{0}{0}a'b'$ is an ideal of $\mathbf{B}$.) Using "$\approx$" to indicate the relation of isomorphism between elements of $\mathbf{B}^*([a] \cup \mathbf{B}^*)^{-1}$ and $\mathbf{B}^*/([a]')$ we have by Theorem 2

$$\left(\frac{[b]}{[a]}\right)_{[a]} \approx [ab] + \frac{1}{0}a'b$$

or, in fuller notation,

$$\left(\frac{[b]}{[a]}\right)_a \approx 1 \cdot ab + 0 \cdot ab' + \frac{0}{0}a'b' + \frac{1}{0}a'b \tag{12}$$

which is what for us corresponds to Boole's

$$\frac{b}{a} = 1 \cdot ab + 0 \cdot ab' + \frac{0}{0}a'b' + \frac{1}{0}a'b.$$

Since he also has $x = b/a$, he can then write

$$x = 1 \cdot ab + 0 \cdot ab' + \frac{0}{0}a'b' + \frac{1}{0}a'b,$$

whereas we can only equate the isomorphic images of the members of (11) to obtain

$$\begin{aligned}[x] + \frac{1}{0}a'b &= [ab] + \frac{1}{0}a'b \\ &= 1 \cdot ab + 0 \cdot ab' + \frac{0}{0}a'b' + \frac{1}{0}a'b. \end{aligned} \tag{13}$$

Now for the Boolean interpretation (i.e., in terms of $\mathbf{B}$) of (13). If $ax = b$ then $a'b = a'ax = 0$, so that

$$\frac{1}{0}a'b = ([a'b]) = ([0])$$

is the zero ideal of $\mathbf{B}^*$. Thus from (13)

$$[x] = [ab] = 1 \cdot ab + 0 \cdot ab' + \frac{0}{0}a'b'$$

and so by (5),

$$x = ab + va'b' \qquad v \in \mathbf{B}.$$

Consequently (7) implies, through (13), that

$$\begin{cases} x = ab + va'b' & v \in \mathbf{B} \\ a'b = 0 \end{cases} \tag{14}$$

which is precisely what Boole gives as his interpretation for

$$x = 1 \cdot ab + 0 \cdot ab' + \frac{0}{0} a'b' + \frac{1}{0} a'b.$$

Boole didn't consider a converse argument (i.e., from (14) to (7)) since it was unnecessary from his point of view. For him the processes of "common" algebra were applicable to logical equations and hence b/a would obviously be the correct solution to $ax = b$. The only problem for him was to obtain the logical meaning of $x = b/a$.

Summary

We have briefly described the nascent abstract algebra ideas within which Boole originated his algebra of logic. While he never made his algebra fully explicit, we inferred that what he did use was, if clarified, a commutative ring with unit, without nilpotents, and having idempotents which stood for classes. By thus hewing closely to "common" algebra Boole could use familiar procedures and techniques. He did not realize that class calculus needed only the idempotents and operations closed with respect to them (i.e., Boolean algebra). Instead he used the ring operations and, in particular, its addition, which is not closed with respect to idempotency. Boole also freely used division to solve equations, introducing then a special, not clearly explained or justified, method for extracting the logical content of the resulting quotient. To explain this we found it necessary to introduce additional structures—factor rings, and rings of quotients for such rings.

We have alluded to the possibility of applications for Boole's algebra—to multisets and to pseudo-Boolean functions. As an additional possibility we would also like to mention that in his *Laws of Thought* Boole uses his peculiar expansions with $1/0$ to present an original approach to conditional probability [**2**, p. 195].

References and Readings

For matters relating to the history of algebra relevant to this paper, we recommend [**5**] and [**7**], and for the history of logic, [**5**]. For an alternative to our justification of Boole's methods, one may consult [**11**, pp. 191–197].

[1] G. Boole, An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities, Walton and Maberley, 1854. Reprinted by Open Court Publ. Co., 1916, 1940, and by Dover, New York, 1951.

[2] T. Hailperin, Boole's Logic and Probability, North-Holland, New York, 1976.

[3] P. L. Hammer (Ivănescu) and S. Rudeanu, Boolean Methods in Operations Research and Related Areas, Springer-Verlag, New York, 1968.

[4] W. S. Jevons, Pure Logic and Other Minor Works, Macmillan, New York, 1890.

[5] M. Kline, Mathematical Thought from Ancient to Modern Times, Oxford Univ. Press, New York, 1972, Chapter 32.

[6] W. Kneale and M. Kneale, The Development of Logic, Oxford Univ. Press, New York, 1962.

[7] E. Koppelman, The calculus of operations and the rise of abstract algebra, Arch. Hist. Exact Sci., 8 (1971) 155–242.

[8] N. H. McCoy, Rings and Ideals, Carus Mathematical Monograph No. 18, The Mathematical Association of America, Washington, D. C., 1948.

[9] G. Peacock, A treatise on algebra, 1842; reprinted, Scripta Math., Yeshiva College, 1940.

[10] R. Sikorski, Boolean Algebras, 3rd ed., Springer, New York, 1969.

[11] N. I. Styazhkin, History of Mathematical Logic from Leibniz to Peano, The M.I.T. Press, 1969 (Originally published as Stanovleniye idey mathematischeskoy logiki, Nauka, 1964).

[12] A. I. Uzkov, On rings of quotients of commutative rings (Russian), Mathematicheskii sbornik (N.S.), 22(64) (1948) 439–441. (Translated as American Mathematical Society Transl. no. 3, 1–5, American Mathematical Society, 1949.)

On Functions Whose Inverse Is Their Reciprocal

RUSSELL EULER
JAMES FORAN
University of Missouri—Kansas City
Kansas City, MO 64110

In the first encounter with functional notation it is not unusual for a student to assume that $f^{-1}(x)$ and $1/f(x)$ have the same meaning. After all, in the algebra of real numbers, the symbols a^{-1} and $1/a$ are used interchangeably. Usually a simple example (almost any familiar monotonic function suffices) will convince the student that the inverse f^{-1} and the reciprocal $1/f$ of a function f are different functions. It's interesting to ask when these two notations *can* be used interchangeably. In this note, we find that functions for which confusion of f^{-1} and $1/f$ makes no difference are not ones likely to be encountered by a student in elementary calculus.

Throughout the discussion that follows, we examine a function f which satisfies the following assumption.

ASSUMPTION. *The function f is one-to-one from the positive half-line $(0,\infty)$ onto itself and satisfies $f^{-1}(x)=1/f(x)$ for all x in $(0,\infty)$.*

We choose the interval $(0,\infty)$ as the domain of f because it and $(-\infty,0)$ are the largest real intervals on which f, its inverse, and its reciprocal can all be defined.

What can be said about the continuity of f? If f is continuous on $(0,\infty)$, then the inverse of f is strictly monotone in the same direction as f (i.e., increasing if f is increasing, and decreasing if f is decreasing), but the reciprocal of f is strictly monotone in the opposite direction. This contradiction shows that f has at least one discontinuity in $(0,\infty)$.

We can often learn about the graph of a function by trying to discover transformations of the plane which leave the graph invariant. The special property of our function f yields the equation $x=f^{-1}(f(x))=1/f(f(x))$, so that $f(f(x))=1/x$ for all $x\in(0,\infty)$. If $a\in(0,\infty)$ and $f(a)=b$, then

$$f(f(a))=f(b)=\frac{1}{a},\qquad f(f(b))=f\left(\frac{1}{a}\right)=\frac{1}{b},$$

and

$$f\left(f\left(\frac{1}{a}\right)\right)=f\left(\frac{1}{b}\right)=a.$$

Thus, if the point (a,b) is on the graph of f, so are the three other points

$$\left(b,\frac{1}{a}\right),\ \left(\frac{1}{a},\frac{1}{b}\right),\text{ and }\left(\frac{1}{b},a\right).$$

Another way of saying this is: if P^* denotes the plane with the coordinate axes removed, then the graph of f is invariant under the transformation of order 4 of P^* defined by $(x,y)\to(y,1/x)$.

We can now determine that $f(1)=1$. For if $(1,b)$ is on the graph of f, so is $(1,1/b)$ so that $b=1/b$ and thus $b=1$. In fact, if $a\in(0,\infty)$ and $f(a)=a$ or $f(a)=1/a$, then $a=1$. For if (a,a) is on the graph of f, so is $(a,1/a)$ and $a=1/a$ implies $a=1$. Similarly, if $(a,1/a)$ is on the graph of f, so is (a,a) and again $a=1$. From this we observe that, *for all* $a\in(0,\infty)$, *except for* $a=1$, *the four points*

$$(a,b),\ \left(b,\frac{1}{a}\right),\ \left(\frac{1}{a},\frac{1}{b}\right),\text{ and }\left(\frac{1}{b},a\right) \tag{1}$$

are distinct points on the graph of f with distinct x-coordinates and distinct y-coordinates.

Using what is technically referred to as the axiom of choice, it is possible to choose quadruples to construct the graphs of functions which satisfy our assumption. But such a construction can yield a function which is discontinuous at every point. We will see that a function satisfying our assumption need not be that badly behaved.

If we remove the curves $x=1$, $y=1$, $y=x$ and $y=1/x$ from the first quadrant of the plane, eight regions remain (see FIGURE 1). It is simple to check that each quadruple of points (1) on the graph of f consists of one point in each of four regions, either the four odd-numbered regions or the four even-numbered regions, and that, except for the point $(1,1)$, no two of these four regions have a common boundary point. Having observed this, we are ready to prove a theorem that shows that f cannot be too nicely behaved.

THEOREM. *A function satisfying our assumption must have infinitely many discontinuities.*

The proof of this theorem involves two possible cases. First suppose that f is continuous at $x=1$. Let $\{x_n\}_{n=1}^{\infty}$ be a strictly increasing sequence of positive real numbers with limit 1, and let $y_n=f(x_n)$. Then the sequences of points $\{(x_n,y_n)\}$ and $\{(y_n,1/x_n)\}$ are on the graph of f, and both must approach the point $(1,1)$. If f had only a finite number of discontinuities, then f would be continuous on an interval I having 1 as right endpoint. Within I, the sequence $\{(x_n,y_n)\}$ must be in region VII or VIII, and $\{(y_n,1/x_n)\}$ must be in region V or VI, so the line $y=1$ separates the points in the two sequences. By the intermediate value theorem, f takes on the value 1 within I, a clear contradiction to the monotonicity of f. Thus, in the case where f is continuous at 1, f has infinitely many discontinuities.

On the other hand, suppose that f is discontinuous at 1 and that f has exactly n discontinuities. Then, since f has an inverse, each discontinuity must be a jump discontinuity and the graph of f must consist of finitely many components each of which is either a single point or the image of an interval of the form $[a,b)$, (a,b), $(a,b]$, or $[a,b]$. If we consider the single points and the images

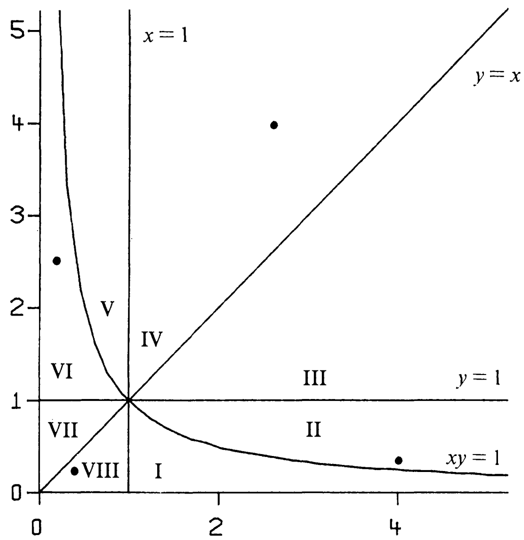


FIGURE 1. **The eight regions with one quadruple of points (1) shown.**

of endpoints of intervals separately, then, because the graph of f is made up of quadruples of points along with the point $(1,1)$, the graph of f consists of $4m+1$ points along with the image of $4n$ open intervals. Consequently, the interval $(0,\infty)$ consists of $4n$ open intervals and $4m+1$ points. Since there must be a point between each consecutive pair of intervals, it follows that there are $4n-1$ points. Then $4n-1=4m+1$. Since this is not possible, we see that, in the event that f is discontinuous at 1, f must also have infinitely many discontinuities.

Let us look at an example of a function which satisfies our assumption.

EXAMPLE 1. (See FIGURE 2.) This function is defined piece-wise on intervals whose endpoints are consecutive integers or their reciprocals. In the definition, n denotes a positive integer.

$$f(x) = \begin{cases} 1 & \text{if } x=1 \\ x-1 & \text{if } x\in(2n,2n+1] \\ 1/(x+1) & \text{if } x\in(2n-1,2n] \\ x/(1-x) & \text{if } x\in[1/(2n+1),1/(2n)] \\ (x+1)/x & \text{if } x\in[1/(2n),1/(2n-1)]. \end{cases}$$

For this function, it is easily observed that $f^{-1}(x)=1/f(x)$ and that f has discontinuities at precisely the points n and $1/n$ for each positive integer n.

Example 1 shows that the discontinuities of a function f satisfying our assumption need not have a limit point in $(0,\infty)$. A second example, given below, shows that it is possible that f be continuous on an interval of the form (M,∞), where $M>1$. Of course in this case the discontinuities of f must have a limit point. More interesting, though, is the fact that in this case f must have a nonzero asymptote.

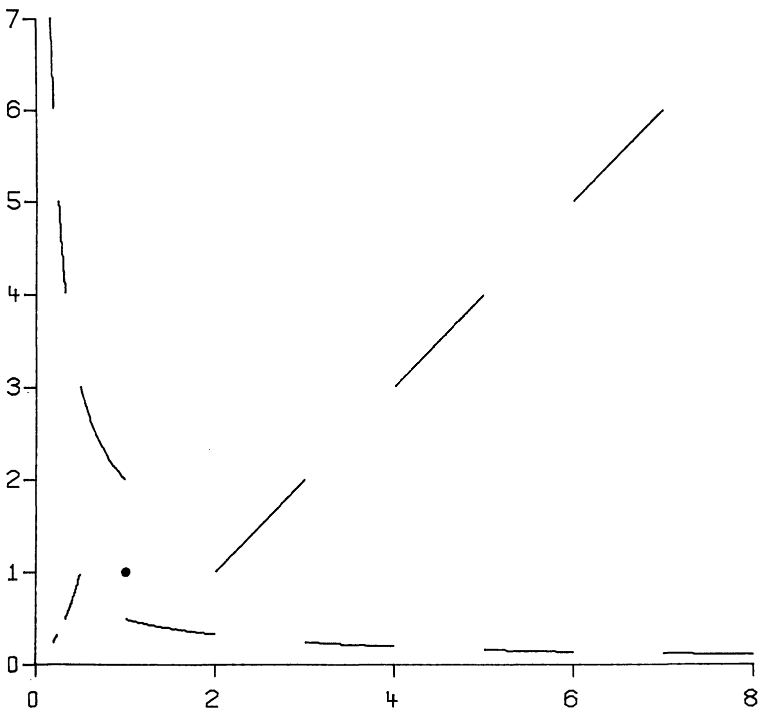


FIGURE 2. **The function given in Example 1.**

To see this, we first show that if f is continuous on (M,∞) where $M>1$, then f must be bounded by M on (M,∞). For suppose (a,b) is on the graph of f where $a>M$ and $b>M$. Then (a,b) is in region III or IV of FIGURE 1 so that $(b,1/a)$ is in region I or II. By the intermediate value theorem, there is an x between a and b such that $f(x)=1$, which is impossible. Since f is continuous on (M,∞), it must be monotone, and since we have just shown that f is bounded by M, f must have a horizontal asymptote. We now suppose that $f(x)\to 0$ as $x\to\infty$ and show that this is not possible (even if f is not continuous on an interval of the form (M,∞)). To see this, let $N>1>f(N)$ be given. Choose a point (a,b) on the graph of f so that $a>N$ and $b<1/N$. Then (a,b) is in region I or II so that $(1/b,a)$ is in region III or IV. Since $1/b>N$, it follows that $y=0$ cannot be an asymptote for any f satisfying our assumption.

Here is an example of a function which satisfies our assumption and is continuous on $(3,\infty)$.

Example 2. (See FIGURE 3.) As in Example 1, the function f is defined piece-wise on intervals. First, we define

$$f(x) = \begin{cases} 1 & \text{if} \quad x=1 \\ 2x/(x+2) & \text{if} \quad x\in[3,\infty) \\ (2-x)/(2x) & \text{if} \quad x\in[\frac{6}{5},2) \\ x+1/2 & \text{if} \quad x\in(0,\frac{1}{3}] \\ 2/(2x-1) & \text{if} \quad x\in(\frac{1}{2},\frac{5}{6}], \end{cases}$$

then complete the definition of $f(x)$ as follows. Define f from $[2,3)$ onto $(1,\frac{6}{5})$ by letting f take each interval $[3-2^{-n+1},3-2^{-n})$ onto $[1+5^{-1}\cdot 2^{-n},1+5^{-1}\cdot 2^{-n+1})$ linearly. The graph of f is completed on the intervals $(\frac{1}{3},\frac{1}{2}]$, $(\frac{5}{6},1)$, and $(1,\frac{6}{5})$ using the fact that $f^{-1}(x)=1/f(x)$. It is easily seen that this f is continuous on $[3,\infty)$ and satisfies $f^{-1}(x)=1/f(x)$ for all $x\in(0,\infty)$.

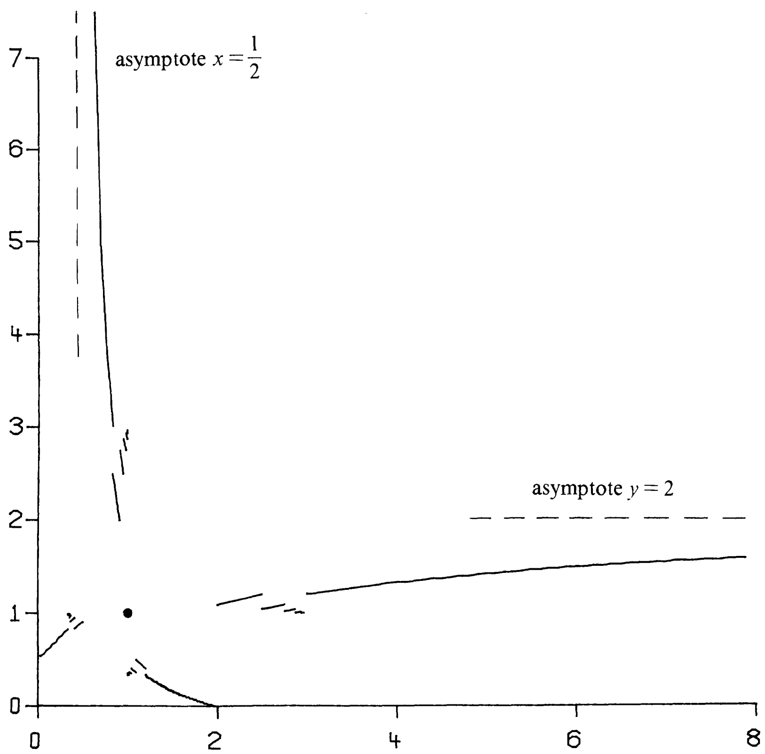


FIGURE 3. **The function given in Example 2.**

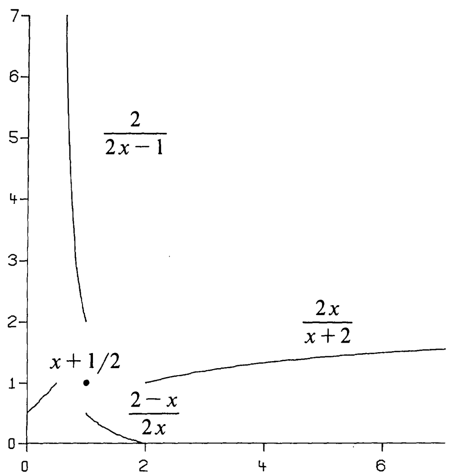


FIGURE 4 (a). **The function in Example 3 with $M=2$.**

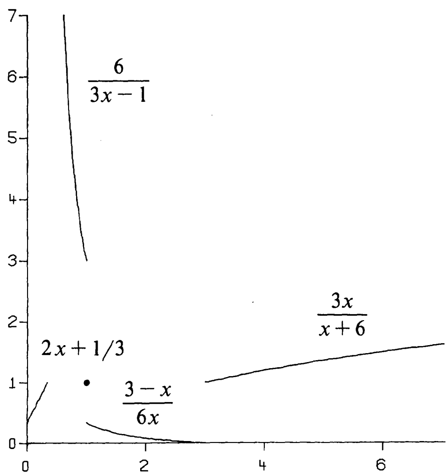


FIGURE 4 (b). **The function in Example 3 with $M=3$.**

Finally, we note that if we extend the domain of our function and allow 0 and ∞ in the domain and range of f (with the convention that $1/0=\infty$ and $1/\infty=0$) then the addition of the two points 0 and ∞ causes the proof of our theorem to break down (because $4m-1$ can equal $4n-1$). Fix $M>1$ and consider the following example.

EXAMPLE 3. (See FIGURE 4.) Let

$$f(x) = \begin{cases} 1 & \text{if} \quad x=1 \\ Mx/(x+M^2-M) & \text{if} \quad x\in(M,\infty] \\ (M-x)/(M^2-M)x & \text{if} \quad x\in(1,M] \\ \left[(M^2-M)x+1\right]/M & \text{if} \quad x\in[0,1/M) \\ (M^2-M)/(Mx-1) & \text{if} \quad x\in[1/M,1). \end{cases}$$

This f satisfies $f^{-1}(x)=1/f(x)$ but has only three discontinuities. Note, however, that the graph of f has five components and, since the graph of f must inhabit four regions of FIGURE 1, the three discontinuities and five components are best possible.

The reader interested in functional equations such as the one considered in this note might find interesting the equations discussed in the papers listed below.

References

[1] F. J. Budder, Functions which permute the roots of an equation, Math. Gaz., 60, 411 (1976) 24–37.

[2] M. J. Hoffman, Continuity of inverse functions, this MAGAZINE, 48, 2 (1975) 66–73.

[3] D. J. McCarthy, Decompositions of functions—functional nth roots of unity, Math. Gaz., 64, 428 (1980) 107–115.

[4] K. Rogers, Solving an exponential equation, this MAGAZINE, 153, 1 (1980) 26–28.

[5] H. Shapiro, Powers of x^2+1, this MAGAZINE, 50, 4 (1977) 208–209.

Draw-A-String: A Slot Machine Game

BRUCE R. JOHNSON
H. PAUL SMITH
University of Victoria
Victoria, British Columbia, Canada V8W 2Y2

Standard mechanically-operated slot machines obey the probability laws of random selection *with* replacement. In contrast, Draw-A-String is an electronic game that operates on the principle of random selection *without* replacement.

Consider the One-Arm Bandit pictured in FIGURE 1 whose operation is based on the probability distribution of the length of the longest string of consecutive integers. Draw-A-String is designed to provide audio and video, as well as monetary, entertainment. The machine will select sequentially at random without replacement from the 36 available numbers. Chosen numbers will light up one at a time until the selection is complete. As each selected number lights up, an appropriate tone will signal the level of contribution that the current selection has made. Upon completion of the selection process the proper payoff line will light up, and the dollar amount will flash on and off, accompanied by musical tones which describe the level of payoff in an entertaining way.

Draw-A-String possesses the following important features:

(a) There are several outcomes for which the player will win something, but only one losing outcome. Also, there is a small chance to win a *large* amount of money!

(b) Player tension builds throughout the selection process with each successive draw having the potential to improve the final outcome.

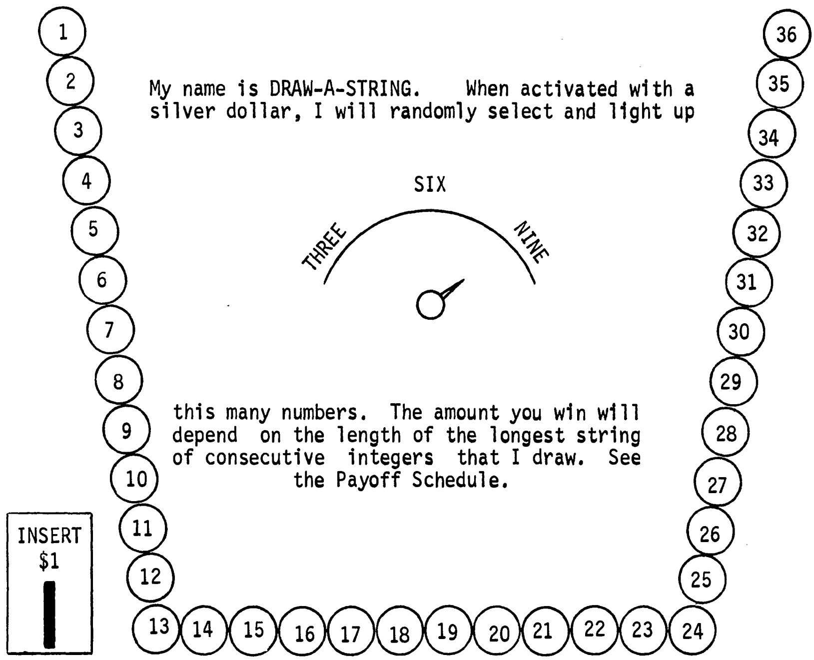
(c) The payoff schedule has been set such that no matter which of the three options is chosen, the persistent gambler will lose an average of about 7 cents per play. Not bad odds, considering that every casino requires a negative expected gain for its slot machines.

Draw-A-String has all the necessary ingredients to be an exciting slot machine game. Plus, you can easily build your own personalized model if you have access to a microcomputer, such as Apple II. While sharpening your programming skills, you will be demonstrating and reinforcing some interesting ideas in combinatorial probability. As a bonus you will have an educational toy to entertain your friends.

We hope you are surprised and somewhat skeptical that such a generous payoff schedule could possibly favor the house at the average rate of 7 cents per play. Proof lies in the derivations that follow. Draw-A-String probabilities and expected payoffs are listed in TABLE 1.

Suppose k integers are to be drawn sequentially at random without replacement from the first n integers, then arranged in increasing order and partitioned into **strings**—maximal length sequences of consecutive integers. (For example, if $n=36$, $k=7$ and the integers 9, 23, 3, 8, 22, 15, 10 were to be drawn, there would be four strings (3; 8-9-10; 15; 22-23) of lengths 1, 3, 1 and 2, respectively.) The probability distribution of the number of strings and related distributions have been investigated extensively in connection with the theory of runs in a series of n elements of which k are indistinguishable a's and $n-k$ are indistinguishable b's. (Of course, forming an arrangement in a row of a's and b's amounts to selecting k positions from $\{1,2,\ldots,n\}$ to be occupied by the a's.) See Bradley [**2**], chapters 11 and 12, for an excellent account of the theory of runs applied to distribution-free statistical tests, including additional references.

We first investigate the probability distribution of the *number of strings*. Let $S_{n,k}$ be the random variable denoting the number of strings that will occur when k integers are drawn at random without replacement from $\{1,2,\ldots,n\}$. According to Mood [**5**] the distribution of $S_{n,k}$ was published first by Ising [**4**] in 1925. To obtain the probability distribution for $S_{n,k}$, a formula that



INSTRUCTIONS

PICK how many numbers you want the machine to select. Your choices are THREE, SIX, or NINE. Study the Payoff Schedule before making your decision.

PLUG one silver dollar into the slot.

PULL the arm.

PRAY that a single string of consecutive integers will light up. There are several other profitable outcomes, however.

PAYOFF SCHEDULE

LENGTH OF LONGEST STRING	HOW MANY NUMBERS ARE SELECTED		
	THREE	SIX	NINE
1	$ 0	$ 1	$ 2
2	$ 3	$ 0	$ 0
3	$ 96	$ 2	$ 1
4		$ 20	$ 2
5		$ 180	$ 10
6		$ 10,000	$ 75
7			$ 1,000
8			$ 10,000
9			$ 250,000

FIGURE 1

counts the number of distinct, equally likely ways to obtain exactly s strings will be derived. We make the selection of k integers from $\{1,2,\dots,n\}$ such that exactly s strings occur in two stages. In stage one, we select the ordered string lengths for the s strings. Each possible selection can be represented uniquely by the arrangement in a row of k indistinguishable stars with $s-1$ indistinguishable bars positioned in $s-1$ of the $k-1$ spaces between the stars. (For example, if $k=7$ and $s=4$, the pattern $*|***|*|**$ would denote 4 strings, with first-order string length = 1, second-order string length = 3, third-order string length = 1 and fourth-order string length = 2.) From the star-bar identification it is obvious that there are exactly $\binom{k-1}{s-1}$ distinct ways to complete stage one. In stage two, we select s positions for the strings from the $n-k+1$ spaces among the $n-k$ ordered (by rank) numbers that will remain behind. (Continuing with the example, if $n=36$, $k=7$, $s=4$, the set of ordered string lengths chosen at stage one is identified by $*|***|*|**$, and if the four positions chosen at stage two are positions 3, 7, 11 and 17, then we have the following scheme

3 8-9-10 15 22-23

(1) (2) (4) (5) (6) (7) (11) (12) (13) (14) (16) (17) (18) (19) (20) (21) (24) (25) ⋯ (36)

positions ∧ 3 ∧ 7 ∧ 11 ∧ 17

which results in the unique selection of $k=7$ integers from $\{1,2,\dots,36\}$ such that $s=4$ strings occur. The seven integers selected are 3,8,9,10,15,22,23.) Stage two can be completed in exactly $\binom{n-k+1}{s}$ different ways. Hence, the multiplication principle implies there are exactly $\binom{k-1}{s-1}\binom{n-k+1}{s}$ distinct ways to choose k integers from $\{1,2,\dots,n\}$ such that s strings occur.

Therefore,

$$P(S_{n,k}=s)=\frac{\binom{k-1}{s-1}\binom{n-k+1}{s}}{\binom{n}{k}},\ s=1,2,\dots,k. \tag{1}$$

Now we can investigate the distribution of the *length of the longest string*. Let $M_{n,k}$ denote the length of the longest string that will occur when k integers are drawn at random without replacement from $\{1,2,\dots,n\}$. Although the probability distribution of $M_{n,k}$ is known (see Bateman [**1**], and Burr and Cane [**3**]), we believe the recursive formula given below is new.

For $1\leqslant m\leqslant k$,

$$P(M_{n,k}=m)=\sum_{s=1}^{k}P(S_{n,k}=s)P(M_{n,k}=m|S_{n,k}=s).$$

Also, given $S_{n,k}=s$, each of the $\binom{k-1}{s-1}$ possible ordered string lengths is equally likely to occur. (It might be helpful to review the derivation of formula (1) at this point.) Therefore,

$$\begin{aligned}P(M_{n,k}=m)&=\sum_{s=1}^{k}P(S_{n,k}=s)\frac{d(k,s,m)}{\binom{k-1}{s-1}}\\&=\frac{1}{\binom{n}{k}}\sum_{s=1}^{k}\binom{n-k+1}{s}d(k,s,m),\end{aligned} \tag{2}$$

where $d(k,s,m)$ is the number of different ways to select s ordered string lengths such that the sum of the string lengths equals k and the longest string length equals m. (Stated another way,

$d(k,s,m)$ is the number of different ways to partition k indistinguishable objects into s ordered, nonempty groups such that the size of the largest group is m.)

Unfortunately, there does not exist a simple combinatorial formula representation for $d(k,s,m)$. However, a recursive formula will be given. For this it is necessary (and reasonable) to define

$$d(0,0,0)=1. \tag{3}$$

Clearly,

$$d(k,s,m)=0 \quad \text{whenever} \quad k<s \quad \text{or} \quad k<m, \tag{4}$$

and

$$d(1,s,m)=\begin{cases} 1 & \text{if } s=1=m, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

If r is the multiplicity of longest string lengths among s ordered string lengths, there are $\binom{s}{r}$ distinct positions to locate the r longest string lengths, and for each of these there are $\sum_{j=0}^{m-1} d(k-rm, s-r, j)$ different ways to order the remaining $s-r$ string lengths into the remaining $s-r$ positions. Thus, it is apparent that, for $1 \leqslant m \leqslant k$,

$$d(k,s,m)=\sum_{r=1}^{[k/m]} \binom{s}{r} \sum_{j=0}^{m-1} d(k-rm, s-r, j). \tag{6}$$

Notice that, because of (3), formula (6) remains valid when m divides k.

Tables of $d(k,s,m)$ are given below for $1 \leqslant s \leqslant k$, $1 \leqslant m \leqslant k$, and $1 \leqslant k \leqslant 9$.

$k=1$

$s \backslash m$	1
1	1

$k=2$

$s \backslash m$	1	2
1	0	1
2	1	0

$k=3$

$s \backslash m$	1	2	3
1	0	0	1
2	0	2	0
3	1	0	0

$k=4$

$s \backslash m$	1	2	3	4
1	0	0	0	1
2	0	1	2	0
3	0	3	0	0
4	1	0	0	0

$k=5$

$s \backslash m$	1	2	3	4	5
1	0	0	0	0	1
2	0	0	2	2	0
3	0	3	3	0	0
4	0	4	0	0	0
5	1	0	0	0	0

$k=6$

$s \backslash m$	1	2	3	4	5	6
1	0	0	0	0	0	1
2	0	0	1	2	2	0
3	0	1	6	3	0	0
4	0	6	4	0	0	0
5	0	5	0	0	0	0
6	1	0	0	0	0	0

$k=7$

$s \backslash m$	1	2	3	4	5	6	7
1	0	0	0	0	0	0	1
2	0	0	0	2	2	2	0
3	0	0	6	6	3	0	0
4	0	4	12	4	0	0	0
5	0	10	5	0	0	0	0
6	0	6	0	0	0	0	0
7	1	0	0	0	0	0	0

$k=8$

$s \backslash m$	1	2	3	4	5	6	7	8
1	0	0	0	0	0	0	0	1
2	0	0	0	1	2	2	2	0
3	0	0	3	9	6	3	0	0
4	0	1	18	12	4	0	0	0
5	0	10	20	5	0	0	0	0
6	0	15	6	0	0	0	0	0
7	0	7	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0

$k=9$

$s \backslash m$	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	1
2	0	0	0	0	2	2	2	2	0
3	0	0	1	9	9	6	3	0	0
4	0	0	16	24	12	4	0	0	0
5	0	5	40	20	5	0	0	0	0
6	0	20	30	6	0	0	0	0	0
7	0	21	7	0	0	0	0	0	0
8	0	8	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0

We conclude by applying formula (2) to obtain the probabilities for Draw-A-String. For example,

$$P(M_{36,9}=2)=\binom{36}{9}^{-1}\sum_{s=1}^{9}\binom{28}{s}d(9,s,2)$$

$$=\binom{36}{9}^{-1}\left\{\binom{28}{5}\cdot 5+\binom{28}{6}\cdot 20+\binom{28}{7}\cdot 21+\binom{28}{8}\cdot 8\right\}$$

$$=\frac{57{,}755{,}880}{94{,}143{,}280}.$$

The complete list of Draw-A-String probabilities is given in TABLE 1.

Length of Longest String	How Many Numbers Are Selected: Three	Six	Nine
1	$\frac{5984}{7140}$	$\frac{736{,}281}{1{,}947{,}792}$	$\frac{6{,}906{,}900}{94{,}143{,}280}$
2	$\frac{1122}{7140}$	$\frac{1{,}042{,}840}{1{,}947{,}792}$	$\frac{57{,}755{,}880}{94{,}143{,}280}$
3	$\frac{34}{7140}$	$\frac{153{,}295}{1{,}947{,}792}$	$\frac{23{,}852{,}556}{94{,}143{,}280}$
4		$\frac{14{,}415}{1{,}947{,}792}$	$\frac{4{,}746{,}924}{94{,}143{,}280}$
5		$\frac{930}{1{,}947{,}792}$	$\frac{767{,}340}{94{,}143{,}280}$
6		$\frac{31}{1{,}947{,}792}$	$\frac{102{,}312}{94{,}143{,}280}$
7			$\frac{10{,}584}{94{,}143{,}280}$
8			$\frac{756}{94{,}143{,}280}$
9			$\frac{28}{94{,}143{,}280}$
Expected Payoff	$.93	$.93	$.93

TABLE 1. **Probabilities for Draw-A-String.**

Acknowledgements

Our presentation has been much improved thanks to suggestions by the referees and editor.

References

[1] G. Bateman, On the power function of the longest run as test for randomness in a sequence of alternatives, Biometrika, 35(1948) 97–112.

[2] J. V. Bradley, Distribution-Free Statistical Tests, Chapters 11 and 12, Prentice-Hall, 1968.

[3] E. J. Burr and Gwenda Cane, Longest run of consecutive observations having a specified attribute, Biometrika, 48(1961) 461–465.

[4] E. Ising, Zeit. f. Physik., 1925, vol. 31, p. 253.

[5] A. M. Mood, The distribution theory of runs, Ann. Math. Statist., 11(1940) 367–392.

Casting out Nines Revisited

Peter Hilton
Case Institute of Technology
Cleveland, OH 44106

Jean Pedersen
University of Santa Clara
Santa Clara, CA 95053

The splendid computational check known as "casting out 9's" has fallen into disuse in our schools. This is hardly surprising in view of the current obsession, at the elementary level, with outmoded computational algorithms at the expense of any real mathematical understanding. Indeed it is our experience that most university students have no idea what the phrase "casting out 9's" means. Easily accessible discussions are rare in contemporary literature; however, they may be found in [**1**], [**3**], [**4**], and [**6**]. The history of the topic may be found in [**5**].

Before explaining the significance of casting out nines, let us establish the notation which will be used in this note.

$\mathbb{Z}$	ring of integers
$\mathbb{Q}$	field of rational numbers
$\mathbb{Z}/n$	ring of integers modulo the positive integer n
$[a]_n$	residue class (in $\mathbb{Z}/n$) of the integer a modulo n (abbreviated to $[a]$ if no confusion need be feared.)
$a \equiv b \bmod n$	a is congruent to b modulo n, i.e., n divides $(b-a)$ (abbreviated to $a \equiv b$ if no confusion need be feared.)
P	a family of prime numbers
$\mathbb{Z}_P$	subring of $\mathbb{Q}$, whose elements are represented by fractions a/b with b prime to the numbers in P
$\mathbb{Z}_n$	subring of $\mathbb{Q}$, whose elements are represented by fractions a/b with b prime to n (*not* to be confused with $\mathbb{Z}/n$)

The projection $\phi = \phi_n : \mathbb{Z} \to \mathbb{Z}/n$ from the integers to the integers modulo n, given by

$$\phi(a) = [a]$$

is a homomorphism of unitary rings; that is to say, ϕ preserves addition, multiplication and the unity element. It follows that any true formula F in $\mathbb{Z}$ (for example, $127(86-29) = 7239$) is transformed into a true formula $\phi(F)$ in $\mathbb{Z}/n$ (in our example, if $n = 5$, the formula transforms to $2(1-4) \equiv 4 \bmod 5$). A check on the truth of F is thus provided by examining $\phi(F)$; for $\phi(F)$ is simpler to verify than F and, if $\phi(F)$ is false, so must F be. Of course, it is perfectly possible for F to be false and $\phi(F)$ true (consider the statement $6 = 1$, which is true modulo 5). Thus if $\phi(F)$ is true we gain a degree of belief in the truth of F. Indeed, the relation between F and $\phi(F)$ is analogous to that between a scientific hypothesis and an experiment designed to test the hypothesis. The experiment can prove the hypothesis false, but it can never *prove* it true.

The reason for choosing $n = 9$ for our check is that the function ϕ_9 is particularly easy to calculate—we give details of the procedure later. We may "cast out 11's" with almost equal facility since it is not difficult to calculate ϕ_{11}; this will also be described. The reader may wonder why we don't "cast out 10's" since it is ridiculously easy to calculate ϕ_{10}—we just take the units digit. The reason is that this would not, in practice, constitute an independent test, since we would already have done the appropriate calculation with the units digits in obtaining the formula F in the first place. On the other hand, the calculations involved in verifying $\phi_9(F)$ and $\phi_{11}(F)$ would be quite independent of our original calculations.

When we discovered that the university students we were teaching had never heard of casting out 9's (this first example of a genuine algebraic homomorphism should have been encountered no later than 6th grade in our judgment), we decided to provide the necessary background ourselves. Thinking about the topic, we quickly realized that the method extends to the checking of calculations involving fractions. It is this experience which led to the title and content of this paper—this extension does not seem to occur anywhere in the literature. Before presenting the formal statement of the extension in Theorem 3, we provide some mathematical background to justify this extension. This background material may be omitted by the reader who may (a) regard the extension as obvious, (b) take it on trust, or (c) provide his or her own argument. The reader who proceeds directly from our presentation of the classical casting out of 9's and of casting out 11's to our discussion of Theorem 3 should notice no sudden jump.

Theorem 3 and the discussion following it describe how, and under what precise circumstances, we may cast out 9's or 11's to check a fractional computation. We are not concerned here with justifying a curriculum in which fractional computations figure prominently—that would be a hard task indeed! Our position is that, if the student is condemned to devote a substantial amount of time to developing skill in such computations, then it would be well if such an effort could be rendered useful and, at least in part, enjoyable. We contend that understanding about the transfer and simplification of mathematical structure is useful, since they are very important aspects of mathematical methodology; and that casting out 9's or 11's from fractions, thus replacing the humdrum by the easy and familiar, is fun.

We close with an appendix—which also may be omitted by the less ambitious reader—showing that the extension of casting out 9's or 11's is, in fact, a special case of **localization theory** (see [**2**]), a theory which has come into some prominence in recent years in view of its applications in algebra and topology.

We now describe how to compute $\phi_9:\mathbb{Z}\to\mathbb{Z}/9$ and $\phi_{11}:\mathbb{Z}\to\mathbb{Z}/11$. We start with ϕ_9. Now $10\equiv 1 \bmod 9$ so it follows that, for any positive integer m,

$$10^m\equiv 1 \bmod 9.$$

Thus if k is any positive integer and if $s(k)$ is the sum of the digits of k (expressed in base 10 notation), then

$$k\equiv s(k) \bmod 9.$$

We may iterate the function s, to form $s(k)$, $s(s(k)),\ldots$, until we reach an integer k' with $1\leqslant k'\leqslant 9$. Then $k\equiv k' \bmod 9$, so that

$$\phi_9(k)=\begin{cases}[k']_9 & \text{if } k'\neq 9,\\ 0 & \text{if } k'=9.\end{cases}$$

(We write 0 instead of $[0]_n$, since $[0]_n$ *is* the zero of the ring $\mathbb{Z}/n$.) We complete the calculation of ϕ_9 by observing that $\phi_9(0)=0$ and

$$\phi_9(-k)=-\phi_9(k) \text{ if } k \text{ is positive.}$$

EXAMPLE 1.

$$\phi_9(127)=\phi_9(10)=\phi_9(1)=[1]_9,$$
$$\phi_9(86)=\phi_9(14)=\phi_9(5)=[5]_9,$$
$$\phi_9(29)=\phi_9(11)=\phi_9(2)=[2]_9,$$
$$\phi_9(7239)=\phi_9(21)=\phi_9(3)=[3]_9,$$
$$\phi_9(-29)=-[2]_9=[7]_9, \text{ since } [2]_9+[7]_9=[9]_9=0.$$

The calculation of ϕ_{11} is based on very similar principles. Since $10\equiv -1 \bmod 11$, it follows that, for any positive integer m,

$$10^m\equiv(-1)^m \bmod 11.$$

Thus if k is any positive integer and if $t(k)$ is the alternating sum of the digits of k (expressed in base 10 notation), counting from the right so that the first digit is the units digit and the contribution of the ith digit d_i of k to $t(k)$ is $(-1)^{i-1}d_i$, then

$$k \equiv t(k) \bmod 11.$$

Of course, $t(k)$ may be negative; but, since $\phi_{11}(-k) = -\phi_{11}(k)$, this does not prevent us from iterating the procedure, so that eventually we reach a number k'' with $-9 \leqslant k'' \leqslant 9$ and $k \equiv k'' \bmod 11$. Then

$$\phi_{11}(k) = \begin{cases} [k'']_{11} & \text{if } 0 \leqslant k'' \leqslant 9, \\ [11 + k'']_{11} & \text{if } -9 \leqslant k'' \leqslant -1. \end{cases}$$

EXAMPLE 2.

$$\begin{aligned} \phi_{11}(127) &= \phi_{11}(7-2+1) &&= \phi_{11}(6) &&= [6]_{11}, \\ \phi_{11}(86) &= \phi_{11}(6-8) &&= \phi_{11}(-2) &&= [9]_{11}, \\ \phi_{11}(29) &= \phi_{11}(9-2) &&= \phi_{11}(7) &&= [7]_{11}, \\ \phi_{11}(7239) &= \phi_{11}(9-3+2-7) &&= \phi_{11}(1) &&= [1]_{11}, \\ \phi_{11}(-29) &= -[7]_{11} = [4]_{11}, \text{ since } [7]_{11} + [4]_{11} = [11]_{11} = 0. \end{aligned}$$

We now recall how we may cast out 9's, or 11's, to check a calculation. Suppose it is claimed that

$$127(86-29) = 7239. \tag{1}$$

We check by applying ϕ_9. Thus if (1) is true, it must also be true (see Example 1) that

$$1(5-2) \equiv 3 \bmod 9. \tag{2}$$

Since (2) *is* true, we have gained a degree of belief in (1). Had the claim been made that $127(86-29) = 7249$, then we would have obtained the *false* congruence $1(5-2) \equiv 4 \bmod 9$ by casting out 9's, thus disproving the claim.

We may recheck (1) by applying ϕ_{11}. Thus if (1) is true it must also be true (see Example 2) that

$$6(9-7) \equiv 1 \bmod 11. \tag{3}$$

Since (3) *is* true, we have now even more confidence in the truth of (1).

We now move toward the promised extension of these checks to fractional calculations. Let P be a family of prime numbers and let Q be the complementary family of prime numbers. Let $\overline{Q}$ be the collection of all positive integers which are products of prime numbers in Q; conventionally, $1 \in \overline{Q}$ for any family Q. Thus if P consists of all odd primes, $\overline{Q}$ consists of all powers of 2. We now point out a crucial property of the ring $\mathbb{Z}/n$.

THEOREM 1. *Let P be the family of prime divisors of n and let Q be the complementary family of prime numbers. Then, in the ring $\mathbb{Z}/n$, we have unique division by the numbers in $\overline{Q}$.*

Proof. Let $b \in \overline{Q}$. We must show that the function $[a] \mapsto [ba]$ is a one-one correspondence from $\mathbb{Z}/n$ to itself. We claim that it suffices to show that this function is one-one; for any one-one function from a *finite* set to itself is a one-one correspondence. Thus we must show that if $ba_1 \equiv ba_2 \bmod n$ then $a_1 \equiv a_2 \bmod n$. But if $n \mid b(a_1 - a_2)$ and n is prime to b (as it must be since $b \in \overline{Q}$), then $n \mid (a_1 - a_2)$, so that our theorem is established.

This theorem and our interest in the rings $\mathbb{Z}/n$ suggest that we take a family of prime numbers Q and consider commutative unitary rings S which admit unique division by the numbers in $\overline{Q}$. We call such rings ***P*-local**, where P is the complement of Q. Let us write s/b for the element of S we get by dividing the element $s \in S$ by $b \in \overline{Q}$. *We claim that the usual rules for manipulating fractions apply to the elements s/b of the P-local ring S.* Thus, to be specific, if $s_1, s_2 \in S$ and $b_1, b_2 \in \overline{Q}$, then

$$\frac{s_1}{b_1}+\frac{s_2}{b_2}=\frac{b_2s_1+b_1s_2}{b_1b_2}, \tag{4}$$

$$\frac{s_1}{b_1}\cdot\frac{s_2}{b_2}=\frac{s_1s_2}{b_1b_2}; \tag{5}$$

moreover, for any integer k, with $s\in S, b\in \overline{Q}$,

$$k\left(\frac{s}{b}\right)=\frac{ks}{b}. \tag{6}$$

We will be content to demonstrate (4). We must show that

$$b_1b_2\left(\frac{s_1}{b_1}+\frac{s_2}{b_2}\right)=b_2s_1+b_1s_2.$$

Now

$$b_1b_2\left(\frac{s_1}{b_1}+\frac{s_2}{b_2}\right)=b_1b_2\left(\frac{s_1}{b_1}\right)+b_1b_2\left(\frac{s_2}{b_2}\right)=b_2s_1+b_1s_2$$

since, by definition,

$$b_1\left(\frac{s_1}{b_1}\right)=s_1,\ b_2\left(\frac{s_2}{b_2}\right)=s_2.$$

The significance of these observations for us lies in the following conclusion. Consider the ring $\mathbb{Z}_P$, which is the subring of the field $\mathbb{Q}$ of rational numbers, consisting of those rational numbers expressible as fractions a/b with $b\in\overline{Q}$. We call $\mathbb{Z}_P$ the **ring of *P*-local integers**; it is easy to see that it is a P-local ring. Then we consider a homomorphism $\phi:\mathbb{Z}\to S$, where S is P-local. The following result follows immediately from what has gone before, in particular, formulae (4)–(6).

THEOREM 2. *If S is P-local, then a ring-homomorphism $\phi:\mathbb{Z}\to S$ has a unique extension to a ring-homomorphism $\psi:\mathbb{Z}_P\to S$. In fact, ψ is given by*

$$\psi\left(\frac{a}{b}\right)=\frac{\phi(a)}{b}.$$

We note finally that we may calculate s/b in S as follows. Since S admits unique division by b, there is a unique element u in S with $bu=1$. Then $s/b=us$.

We now return to the case where $S=\mathbb{Z}/n$ and thus to our discussion of casting out 9's and 11's. Whether by means of the preceding arguments or by the reader's own method of proof, we have the following key result.

THEOREM 3. *Let $\mathbb{Z}_n$ be the subring of the field $\mathbb{Q}$ of rational numbers, consisting of those rational numbers expressible as fractions a/b with b prime to n. Then the projection $\phi:\mathbb{Z}\to\mathbb{Z}/n$ has a unique extension to a ring-homomorphism $\psi:\mathbb{Z}_n\to\mathbb{Z}/n$, given by*

$$\psi\left(\frac{a}{b}\right)=\frac{\phi(a)}{b}=\frac{[a]}{b}. \tag{7}$$

We now remark that the symbol $[a]/b$ in (7) means that unique residue class $[c]$ such that $b[c]=[a]$. But then $[b][c]=[a]$ so that, to calculate $[a]/b$, we have only to calculate $\phi(b)=[b]$, and then to find the inverse $[\bar{b}]$ of $[b]$ for $1\leqslant b\leqslant n-1$ (and b prime to n). For then $[a]/b=[a\bar{b}]$. We have already discussed the calculation of $\phi(b)$ for $n=9$ or 11, so we will simply give the table of relevant multiplicative inverses for $n=9$ and for $n=11$.

	Multiplicative inverses in $\mathbb{Z}/9$					
b	1	2	4	5	7	8
$\bar{b}$	1	5	7	2	4	8

	Multiplicative inverses in $\mathbb{Z}/11$									
b	1	2	3	4	5	6	7	8	9	10
$\bar{b}$	1	6	4	3	9	2	8	7	5	10

The reader will easily verify that, in the first table, $b\bar{b}\equiv 1 \bmod 9$; and, in the second table, $b\bar{b}\equiv 1 \bmod 11$.

We are now ready to cast out 9's or 11's in fractional calculations. We begin with casting out 9's. In this case, $\mathbb{Z}_9$ is the same as $\mathbb{Z}_3$, the ring of rational numbers expressible as fractions with denominators prime to 3. We adopt this notation since it is more natural and is, moreover, consistent with that in the discussion before Theorem 2, being the ring of 3-local integers. Thus we suppose we have a calculation involving the addition, subtraction and multiplication of fractions, where no fraction has a denominator divisible by 3. We check by casting out 9's as illustrated in the next two examples, where we adopt an evident extension of the congruence notation.

EXAMPLE 3. Check $5\frac{13}{28}+14\frac{71}{220}=19\frac{303}{385}$. By casting out 9's

$$\begin{aligned}
5\frac{13}{28}&\equiv 5+\frac{4}{1}\equiv 0,\\
14\frac{71}{220}&\equiv 5+\frac{8}{4}=7,\\
19\frac{303}{385}&\equiv 1+\frac{6}{7}\equiv 1+(6\times 4)\equiv 7;\\
0+7&=7.
\end{aligned}$$

EXAMPLE 4. Check $6\frac{12}{29}\times 2\frac{3}{71}=13\frac{7}{71}$. By casting out 9's

$$\begin{aligned}
6\frac{12}{29}&\equiv 6+\frac{3}{2}\equiv 6+6\ \equiv 3,\\
2\frac{3}{71}&\equiv 2+\frac{3}{8}\equiv 2+24\equiv 8,\\
13\frac{7}{71}&\equiv 4+\frac{7}{8}\equiv 4+56\equiv 6;\\
3\times 8&\equiv 6.
\end{aligned}$$

We now consider the technique of casting out 11's. Now we extend ϕ: $\mathbb{Z}\to\mathbb{Z}/11$ to ψ: $\mathbb{Z}_{11}\to\mathbb{Z}/11$. We note a certain advantage to casting out 11's over casting out 9's in a fractional calculation, since we are less likely to encounter a fraction with denominator divisible by 11 than one with denominator divisible by 3—the probability in the first case (for a randomly chosen fraction) is $1/11$ and in the second case $1/3$. (We would argue that for fractions not chosen randomly it is even more unlikely to meet one whose denominator is divisible by 11!)

Now we cannot cast out 11's to check the calculation in Example 3 because we *do* have fractions there whose denominators are divisible by 11. (The best we can do in Example 3 is to check that $14\frac{71}{20}\equiv 19\frac{303}{35}$ mod 11, as the reader should understand.) However, we can check the calculation in Example 4 by casting out 11's as follows:

$$\begin{aligned}
6\frac{12}{29}&\equiv 6+\frac{1}{7}\equiv 6+8\equiv 3,\\
2\frac{3}{71}&\equiv 2+\frac{3}{-6}=2-\frac{1}{2}\equiv 2-6=-4\equiv 7,\\
13\frac{7}{71}&\equiv 2+\frac{7}{-6}\equiv 2-(7\times 2)=-12\equiv 10;\\
3\times 7&\equiv 10.
\end{aligned}$$

We close with an example which is not susceptible to the check by casting out 9's.

EXAMPLE 5. Check $5\frac{4}{21}(7\frac{9}{10}-5\frac{2}{15})=14\frac{227}{630}$. By casting out 11's

$$5\frac{4}{21}\equiv 5-4=1,$$

$$7\frac{9}{10}\equiv 7-9=-2,$$

$$5\frac{2}{15}\equiv 5+\frac{1}{2}\equiv 5+6\equiv 0,$$

$$14\frac{227}{630}\equiv 3+\frac{7}{3}\equiv 3+(7\times 4)=31=-2;$$

$$1\times(-2)=-2.$$

We note that the procedures we have given for computing ϕ and ψ, while algorithmic, are not necessarily the most efficient in any particular case—in this respect, these procedures strongly resemble the familiar algorithms of traditional arithmetic. The alert reader will have realized that we can often compute ϕ and ψ more quickly by "tricks." In this way, for example, we may often avoid appealing to the tables of inverses; thus,

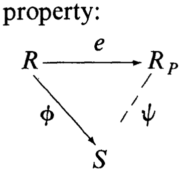
$$\psi_9\left(\frac{51}{71}\right)=\psi_9\left(\frac{51}{17}\right)=[3]_9,$$

$$\psi_9\left(\frac{128}{511}\right)=\psi_9\left(\frac{2}{7}\right)=\psi_9\left(\frac{2}{-2}\right)=\psi_9(-1)=[8]_9,$$

$$\psi_{11}\left(\frac{863}{1329}\right)=\psi_{11}\left(\frac{5}{9}\right)=\psi_{11}\left(\frac{-6}{-2}\right)=[3]_{11}.$$

Appendix: *P*-localization

We mention in this appendix that our extension $\psi:\mathbb{Z}_n\to\mathbb{Z}/n$ of the projection $\phi:\mathbb{Z}\to\mathbb{Z}/n$ is in fact an example of the notion of P-localization (where P is a family of primes) which has been prominent in algebra and homotopy theory in recent years. The discussion following Theorem 1 introduced the concept of a P-local ring S. Now given any commutative unitary ring R we may ask whether it has a **P-localization** R_P. This would be a best possible approximation to R by a P-local ring in the following precise sense. R_P is a P-local ring and there is a ring-homomorphism $e: R\to R_P$ with the following universal property:



Given a homomorphism $\phi: R\to S$ from R to a P-local ring S, there is a unique homomorphism $\psi: R_P\to S$ such that $\psi e=\phi$. Thus what we showed earlier was that $\mathbb{Z}_P$, as there defined, is the P-localization of $\mathbb{Z}$ with $e:\mathbb{Z}\to\mathbb{Z}_P$ being the standard embedding obtained by regarding an integer as a rational number.

The construction of R_P closely imitates the construction of $\mathbb{Z}_P$ out of $\mathbb{Z}$ but, in the general case, we must allow for the possibility that R has torsion prime to P. We say that a ring R (or, indeed, an abelian group R) has **torsion** if there is a nonzero element $a\in R$ and a positive integer n such that $na=0$. For such a **torsion element** a we say that its **order** is k if k is the smallest positive integer n such that $na=0$; and we say that R has **torsion prime to P** if there is a torsion element whose order is prime to all the numbers in the family P. Now it is obvious that a P-local ring cannot have torsion prime to P. For if k is prime to P and $ka=0$ in a P-local ring, then $a=0$. Thus if R has torsion prime to P, then $e: R\to R_P$ cannot be injective!

In fact, to define R_P and $e: R \to R_P$ in the general case, we adopt the following procedure which reduces to the same procedure as that used to construct $\mathbb{Z}_P$ out of $\mathbb{Z}$ provided that R has no torsion prime to P.

We consider pairs (x, m), $x \in R$, $m \in \mathbb{Z}$ and prime to P, and introduce the equivalence relation

$$(x, m) \sim (y, n) \Leftrightarrow \exists k, \text{ prime to } P, \text{ with } knx = kmy.$$

We write x/m for the equivalence class of (x, m), and we add and multiply the "fractions" x/m by the usual rules. Note that we may always take m positive, and that $e: R \to R_P$ is given by $e(x) = x/1$, as usual. However, in the presence of torsion prime to P, we will have cases where $x/1 = y/1$ although $x \neq y$. Indeed, if a has order k prime to P, $k \neq 1$, then $a/1 = 0/1$ but $a \neq 0$.

If $\phi: R \to S$ is a homomorphism from R to the P-local ring S, then $\psi: R_P \to S$ is given, as in the special case $R = \mathbb{Z}$, by

$$\psi\left(\frac{x}{m}\right) = \frac{\phi(x)}{m},$$

where, on the right, we mean by s/m that unique element $t \in S$ such that $mt = s$. It is easy to see that this is the *only* ring-homomorphism such that $\psi(x/1) = \phi(x)$.

References

[1] Roy Dubisch, Basic Concepts of Mathematics for Elementary Teachers, Addison-Wesley, 1977, pp. 424–426.

[2] Peter Hilton, Localization in topology, Amer. Math. Monthly, 82(1975) 113–131.

[3] Peter Hilton and Jean Pedersen, Fear No More: An Adult Approach to Mathematics, Addison-Wesley, 1981, Chapter 6, Section 4.

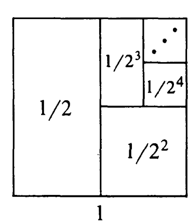
[4] Oystein Ore, Invitation to Number Theory, MAA New Mathematical Library 20, pp. 98–100.

[5] Oystein Ore, Number Theory and Its History, McGraw-Hill, 1948, pp. 225–233.

[6] Edward L. Spitznagel, Jr., Selected Topics in Mathematics, Holt, Rinehart and Winston, 1971, pp. 122–127.

Proof Without Words: Geometric Sums

$$\sum_{k=1}^{n} 1/2^k = 1 - 1/2^n \Rightarrow \sum_{k=1}^{\infty} 1/2^k = 1$$



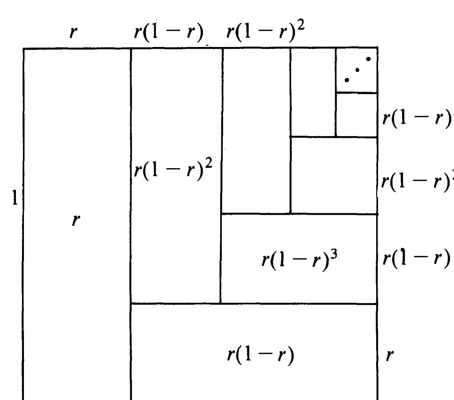
For any $0 < r < 1$,

$$(1-r)^k - r(1-r)^k = (1-r)^{k+1}$$

$$\sum_{k=0}^{n} r(1-r)^k = 1 - (1-r)^{k+1}$$

$$\Downarrow$$

$$\sum_{k=0}^{\infty} (1-r)^k = 1/r$$



—Warren Page
New York City Technical College
Brooklyn, NY 12201

In fact, to define R_P and $e: R \to R_P$ in the general case, we adopt the following procedure which reduces to the same procedure as that used to construct $\mathbb{Z}_P$ out of $\mathbb{Z}$ provided that R has no torsion prime to P.

We consider pairs (x, m), $x \in R$, $m \in \mathbb{Z}$ and prime to P, and introduce the equivalence relation

$$(x,m) \sim (y,n) \Leftrightarrow \exists k, \text{ prime to } P, \text{ with } knx = kmy.$$

We write x/m for the equivalence class of (x, m), and we add and multiply the "fractions" x/m by the usual rules. Note that we may always take m positive, and that $e: R \to R_P$ is given by $e(x) = x/1$, as usual. However, in the presence of torsion prime to P, we will have cases where $x/1 = y/1$ although $x \neq y$. Indeed, if a has order k prime to P, $k \neq 1$, then $a/1 = 0/1$ but $a \neq 0$.

If $\phi: R \to S$ is a homomorphism from R to the P-local ring S, then $\psi: R_P \to S$ is given, as in the special case $R = \mathbb{Z}$, by

$$\psi\left(\frac{x}{m}\right) = \frac{\phi(x)}{m},$$

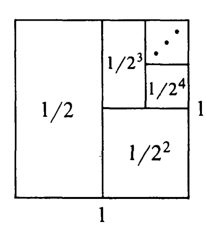
where, on the right, we mean by s/m that unique element $t \in S$ such that $mt = s$. It is easy to see that this is the *only* ring-homomorphism such that $\psi(x/1) = \phi(x)$.

References

[1] Roy Dubisch, Basic Concepts of Mathematics for Elementary Teachers, Addison-Wesley, 1977, pp. 424–426.
[2] Peter Hilton, Localization in topology, Amer. Math. Monthly, 82(1975) 113–131.
[3] Peter Hilton and Jean Pedersen, Fear No More: An Adult Approach to Mathematics, Addison-Wesley, 1981, Chapter 6, Section 4.
[4] Oystein Ore, Invitation to Number Theory, MAA New Mathematical Library 20, pp. 98–100.
[5] Oystein Ore, Number Theory and Its History, McGraw-Hill, 1948, pp. 225–233.
[6] Edward L. Spitznagel, Jr., Selected Topics in Mathematics, Holt, Rinehart and Winston, 1971, pp. 122–127.

Proof Without Words: Geometric Sums

$$\sum_{k=1}^{n} 1/2^k = 1 - 1/2^n \Rightarrow \sum_{k=1}^{\infty} 1/2^k = 1$$



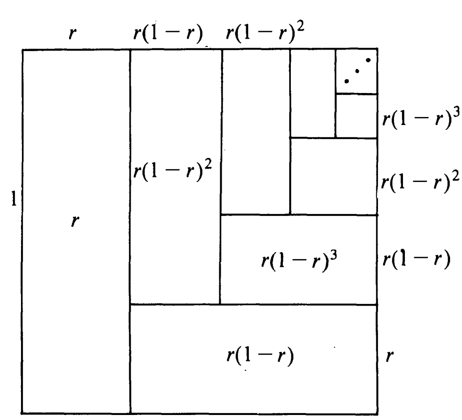
For any $0 < r < 1$,

$$(1-r)^k - r(1-r)^k = (1-r)^{k+1}$$

$$\sum_{k=0}^{n} r(1-r)^k = 1 - (1-r)^{k+1}$$

$$\Downarrow$$

$$\sum_{k=0}^{\infty} (1-r)^k = 1/r$$

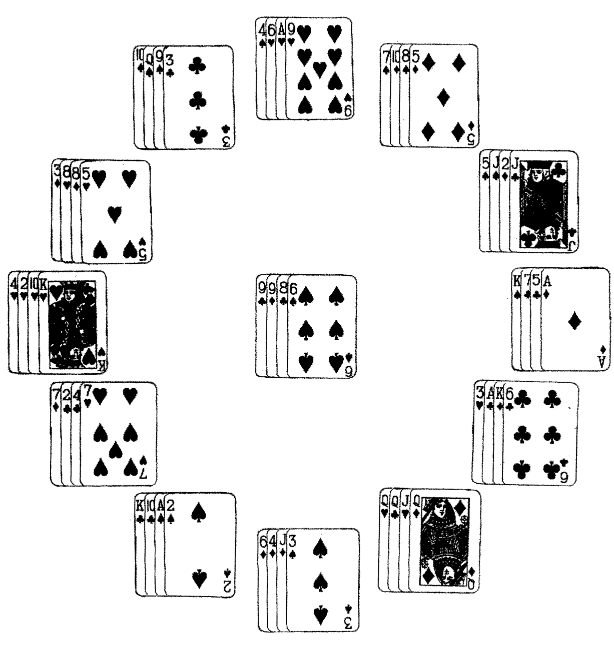


—WARREN PAGE
New York City Technical College
Brooklyn, NY 12201

A Probabilistic Analysis of Clock Solitaire

T. A. Jenkyns
E. R. Muller
Brock University
St. Catharines, Ontario, Canada L2S 3A1

Clock Solitaire is a game in which players have great expectations of winning since the expected number of cards turned over is large (42.4 out of 52) and yet the probability of winning is small (1/13)! Clock Solitaire uses a regular deck of 52 playing cards. (Before reading further, you may wish to see that such a deck is handy for reference and verification of the discussion.) The cards are shuffled and dealt face down in 13 piles of 4 cards each, the piles arranged in the twelve numeral positions of a clock with the thirteenth pile at the centre. To begin the game, the topmost card of the central pile is turned, its face value noted, and then it is placed face up under the pile whose numerical position corresponds to this face value (the values assigned to picture cards are: 1 for Ace, 11 for Jack, 12 for Queen and the centre for King). The topmost unturned card of this pile (under which the first card was placed) is now taken and its face value directs the player to another pile, under which this second card is placed face up. The topmost unturned card of this pile is turned, and the procedure repeated until such time as the fourth King is drawn. There are then no further unturned cards on the central pile. The game is won if there are no unturned cards on any of the other twelve piles; otherwise, the game is lost. A clear description of winning distributions of cards and proof that the probability of a winning game is 1/13 is given by David Kleiner as the solution to Problem 1066, this Magazine [1]. We wish to analyze a "continued" game of Clock, as well as investigate the expected number of cards turned over in a game or its continuation.



A winning game, viewed from below, with cards dealt face down on a glass table. First card drawn is nine of clubs.

We can extend the rules of Clock Solitaire to continue playing until all cards are turned face up and will refer to the first sequence of play, which ends when all Kings are drawn, as the *first play* of the game. If the game is lost, begin a *second play* by taking the top card from the pile in the lowest numerical position (say k) with an unturned card. The second play ends when the player draws the last card with face value k. If some unturned cards still remain in other piles, the game may be continued with a *third play* and so on.

The initial deal determines the order in which all 52 cards are turned face up. In fact every ordering of the cards may be realized as such an **order of play** and we shall assume that all 52! orders are equally likely. The probability of winning and the expected length of the first play depend on the probability $P_1(j)$ of finding the 4th King in a particular position j $(1 \leq j \leq 52)$ in the order of play. Clearly $P_1(1) = P_1(2) = P_1(3) = 0$. The probability $P_1(j)$ can be found from FIGURE 1 where there are 4 ways (Kings) to fill position j, and, since all the Kings appear by position j, there are 48 ways to fill position 52, 47 ways to fill position 51, etc. Hence

$$P_1(j) = \frac{4[(48)(47)\cdots((j+1)-4)](j-1)!}{52!}$$

$$= \frac{4(j-1)(j-2)(j-3)}{(52)(51)(50)(49)} \tag{1}$$

and as a special case P (win in the first play) $= P_1(52) = 1/13$.

# of possible cards to fill position	1	2	...	$j-1$	4	$(j+1)-4$	...	47	48
position in the order of play	1	2	...	$j-1$	j	$j+1$	...	51	52

FIGURE 1

The expected length of the first play is given by

$$\text{EL}_1 = \sum_{j=1}^{52} jP_1(j). \tag{2}$$

This sum can be calculated using the method of differences, that is, if we define

$$F_k(j) = j(j-1)\cdots(j-k),$$

then

$$(k+2)F_k(j) = F_{k+1}(j+1) - F_{k+1}(j)$$

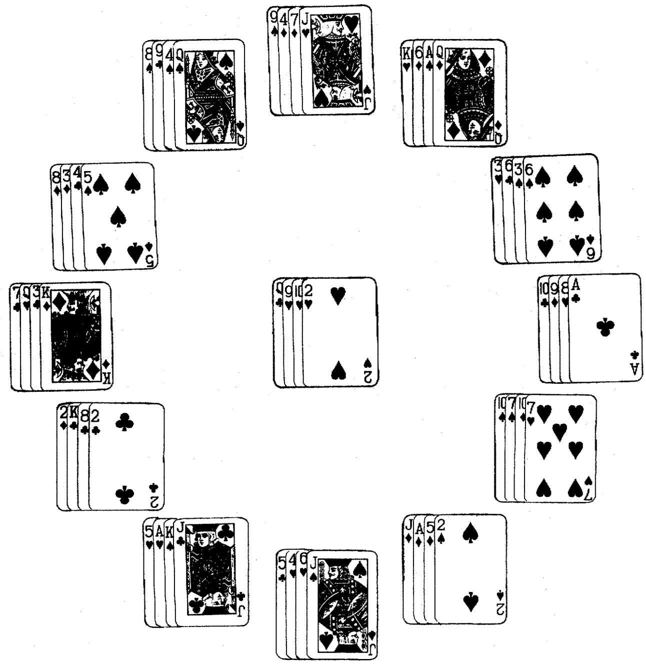
and

$$(k+2)\sum_{j=1}^{n} F_k(j) = \sum_{j=1}^{n} F_{k+1}(j+1) - \sum_{j=1}^{n} F_{k+1}(j)$$
$$= F_{k+1}(n+1) - F_{k+1}(1),$$

as terms cancel out in pairs. Using this method of differences for the summation arising in (2) with $P_1(j)$ given in expression (1), the expected length of the first play is found to be

$$\text{EL}_1 = \frac{4}{(52)(51)(50)(49)} \sum_{j=1}^{52} (j)(j-1)(j-2)(j-3) = 42.4 \text{ cards.}$$

This relatively high expectation value is what makes the game interesting. However, the low probability of winning makes it frustrating!



A lengthy, but losing game.

Suppose that after the first play there are $N=\Sigma_{j=1}^{12} n_j$ unturned cards, where n_j is the number of unturned cards with face value j. (There are no unturned Kings as the first play ended when the 4th King was turned.) Suppose n_k is the first nonzero n_j. To simplify future notation we put $m=n_k$.

The second play is a generalized version of Clock Solitaire with a deck of N cards, n_j cards of denomination j. The game is started by turning the topmost card in pile k, where there are m unturned cards. It is again assumed that all $N!$ orders of play are equally likely. Then the probability $P_2(N,m;j)$ that the last card with face value k is the jth in the order of the second play is obtained from FIGURE 2, i.e.,

$$P_2(N,m;j)=\frac{m}{N}\frac{(N-m)!}{(N-1)!}\frac{(j-1)!}{(j-m)!}. \tag{3}$$

This expression reduces to expression (1) when $N=52$ and $m=4$.

# of possible cards to fill position	1	2	...	$j-1$	m	$j+1-m$	...	$N-1-m$	$N-m$
position in the order of play	1	2	...	$j-1$	j	$j+1$	...	$N-1$	N

FIGURE 2

The expected length of the second play can now be determined, viz.,

$$\mathrm{EL}_2(N,m) = \sum_{j=1}^{N} jP_2(N,m;j) = \frac{m(N+1)}{(m+1)} \tag{4}$$

where once again the sum may be obtained using the method of differences.

We can generalize still further and suppose that the pack of N cards is made up of n different ranks of cards (as opposed to 13 denominations) and m distinguishable cards of each rank (as opposed to 4 suits), i.e., $N = nm$. The cards are dealt in piles of m cards each, placed in positions $1,2,\ldots,n$ and the first play begins by turning the topmost card in pile 1 and continues by following the previously defined rules. Subsequent plays start from the lowest numbered pile which still has an unturned card. It is assumed that all the $N!$ orders of play are equally likely.

If all m cards of each of the ranks 2 to $\alpha - 1 (\alpha \leqslant n)$ have been turned before the last 1 is turned, the second play will start with a card from pile α and will terminate when the mth card of rank α is turned. Thus the number of plays for a given ordering of all N cards is determined by the relative positions of the last card of each rank alone, i.e., by the corresponding permutation of $I_n = \{1,2,\ldots,n\}$. Furthermore, each permutation σ of I_n arises from the same number of orders of play, and so all $n!$ permutations are equally likely. For each permutation σ let $p(\sigma)$ denote the number of plays and be called the *length* of σ. For example, if $n = 8$ the permutation $\sigma = (2,1,6,7,4,3,8,5)$ has length $p(\sigma) = 3$, since the first play ends at 1, the second at 3 and the third at 5.

Let $f(n,p)$ denote the number of permutations of I_n of length p. Then the probability that a game ends after p plays, $P(n,p)$, is given by

$$P(n,p) = \frac{f(n,p)}{n!}. \tag{5}$$

We now explore the quantity $f(n,p)$. A complete listing can be made for small values of n. For example for $n = 3$, the listing

σ	$p(\sigma)$
123	3
132	2
213	2
312	2
231	1
321	1

shows that $f(3,3) = 1$, $f(3,2) = 3$, and $f(3,1) = 2$. In general

$$f(n,p) = 0 \text{ unless } 1 \leqslant p \leqslant n,$$

$$f(n,1) = (n-1)! \tag{6}$$

and

$$f(n,n) = 1.$$

It remains to find a useful expression for $f(n,p)$ when $2 \leqslant p \leqslant n$. One procedure is through the recurrence relation given by the following proposition.

PROPOSITION 1.

$$f(n+1,p+1) = f(n,p) + nf(n,p+1). \tag{7}$$

Proof. Suppose σ is a permutation of I_n. From σ we can generate $n+1$ permutations of I_{n+1} as follows: for $\alpha = 1,2,\ldots,n+1$ let $\sigma(\alpha)$ begin with α, then let the remaining n entries be obtained from σ by leaving them unchanged if they are less than α, and increasing by 1 if they are greater than or equal to α. For instance, if $\sigma = (3,1,2,4)$ then $\sigma(1) = (1,4,2,3,5)$ and $\sigma(3) = (3,4,1,2,5)$. Then each σ of length p generates one permutation, $\sigma(1)$, of length $p+1$ and n permutations of length p. Since no two generated permutations are identical, and since every permutation of I_{n+1} is generated from some permutation of I_n, we have equality (7) as asserted.

Applying induction and (7) it is easy to prove

PROPOSITION 2.

$$\frac{f(n+1,2)}{n!}=1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}. \tag{8}$$

Using the recurrence relation (7) together with the expressions in (6) one can compute $P(n,p)$, the probability that the game ends after p plays. TABLE 1 gives these values for $n=13$ where $Q(13,p)=\Sigma_{j=1}^{p}P(13,j)$.

p	$f(13,p)$	$P(13,p)$	$Q(13,p)$
1	479001600	0.07692	0.07692
2	1486442880	0.23871	0.31563
3	1931559552	0.31019	0.62582
4	1414014888	0.22708	0.85290
5	657206836	0.10554	0.95844
6	206070150	0.03309	0.99153
7	44990231	0.00723	0.99876
8	6926634	0.00111	0.99987
9	749463	0.00012	0.99999
10	55770	0.00001	1.00000
11	2717	0.00000	1.00000
12	78	0.00000	1.00000
13	1	0.00000	1.00000

TABLE 1

The expected number of plays for a game using n ranks of cards is given by

$$E_n=\sum_{p=1}^{n} pP(n,p)=\frac{1}{n!}\sum_{p=1}^{n} pf(n,p). \tag{9}$$

Surprisingly, this expected value can be related to the partial sums of a harmonic series through the following proposition.

PROPOSITION 3.

$$\sum_{p=1}^{n} pP(n,p)=1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}. \tag{10}$$

Proof.

$$\text{For } n=1,\qquad P(1,1)=\frac{f(1,1)}{1!}=1.$$

$$\text{For } n=2,\ \sum_{p=1}^{2} pP(2,p)=\frac{f(2,1)}{2!}+\frac{2f(2,2)}{2!}$$
$$=\frac{1}{2}+1.$$

$$\text{For } n=3,\ \sum_{p=1}^{3} pP(3,p)=\frac{f(3,1)}{3!}+\frac{2f(3,2)}{3!}+\frac{3f(3,3)}{3!}$$
$$=\frac{1}{3}+1+\frac{1}{2}.$$

Assume (10) for $n=r$; then for $n=r+1$

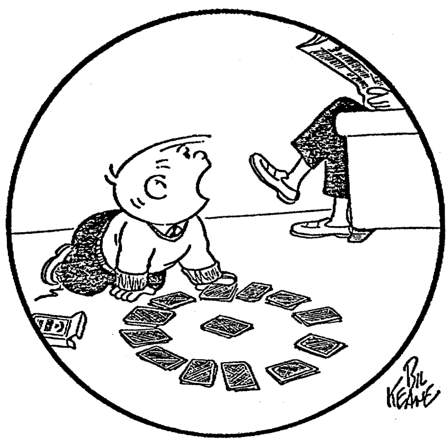
$$\sum_{p=1}^{r+1} pP(r+1,p)=\sum_{p=1}^{r+1} p\frac{f(r+1,p)}{(r+1)!}$$
$$=\frac{1}{(r+1)!}\left[\sum_{p=1}^{r+1} pf(r,p-1)+r\sum_{p=1}^{r} pf(r,p)\right]$$

$$= \frac{1}{(r+1)!}\left[\sum_{q=0}^{r}(1+q)f(r,q) + r\sum_{p=1}^{r} pf(r,p)\right]$$

$$= \frac{1}{r+1} + \frac{r+1}{(r+1)!}\sum_{p=1}^{r} pf(r,p)$$

$$= \frac{1}{r+1} + \left(\frac{1}{r} + \frac{1}{r-1} + \cdots + \frac{1}{2} + 1\right).$$

Using Proposition 3, the expected number of plays for the extended Clock Solitaire with $n = 13$ in expression (10) is $1 + 1/2 + \cdots + 1/13 = 3.18013$. Roughly, this says that a player should expect to "continue" the game for 3 plays in order to finish with all cards turned up.



"Daddy, will you watch me so I don't cheat?"

Earlier, we showed that the expected length of the first play (the number of cards turned over) was $EL_1 = 42.4$ cards. What happens in the extended game? We will close by computing T_q, the expected number of cards turned when the game is stopped as one of the following conditions is met; either (i) all the cards are turned or (ii) q plays have been made. (For $q = 1, EL_1 = T_1$.)

The definition of T_q gives

$$T_q = \sum_{p=1}^{q-1} NP(n,p) + \sum_{j=1}^{N} jP(q\text{th play ends after } j \text{ cards}). \tag{11}$$

The probability, P (qth play ends after j cards), is obtained through the probability $P(q, \alpha, j)$ that the qth play ends after j cards with the last card of rank α. To count the orders of play where the jth card is α and all cards less or equal to α occur among the first j played, one can use FIGURE 3, similar to FIGURES 1 and 2.

# of possible cards to fill position	1	2	...	$j-1$	m	$j+1-m\alpha$	...	$N-1-m\alpha$	$N-m\alpha$
position in the order of play	1	2	...	$j-1$	j	$j+1$	...	$N-1$	N

FIGURE 3

There are

$$\frac{f(\alpha-1, q-1)}{(\alpha-1)!}(j-1)!$$

orderings of the first $(j-1)$ cards which ensure that exactly $(q-1)$ plays have been completed by the $(j-1)$st card and therefore

$$P(q,\alpha,j)=\frac{f(\alpha-1,q-1)}{(\alpha-1)!}\frac{(j-1)!}{(N-1)!}\frac{(N-m\alpha)!}{(j-m\alpha)!}\frac{m}{N}. \tag{12}$$

As a check on equation (12), one can obtain the expression (5) for $P(n,p)$ by summing over α with $j=N$ and $q=p$, whence

$$P(n,p)=\sum_{\alpha=1}^{n}P(p,\alpha,N)=\frac{1}{n}\sum_{\alpha=1}^{n}\frac{f(\alpha-1,p-1)}{(\alpha-1)!}$$
$$=\frac{1}{n}\sum_{\alpha=1}^{n}\left[\frac{f(\alpha,p)-(\alpha-1)f(\alpha-1,p)}{(\alpha-1)!}\right]=\frac{f(n,p)}{n!}$$

as all terms cancel out in pairs and only the last one remains.

For the case $N=52$ and $m=4$ the second term on the right-hand side of expression (11) can now be written and reduced as follows: $\Sigma_{j=1}^{52} jP(q\text{th play ends after } j \text{ cards})$

$$=\sum_{j=1}^{52}j\sum_{\alpha=\left[\frac{j}{4}\right]}^{13}P(q,\alpha,j)$$
$$=\frac{4}{52!}\sum_{\alpha=1}^{13}\frac{(52-4\alpha)!f(\alpha-1,q-1)}{(\alpha-1)!}\sum_{j=4\alpha}^{52}\frac{j!}{(j-4\alpha)!}$$
$$=(4)(53)\sum_{\alpha=1}^{13}\frac{f(\alpha-1,q-1)}{(4\alpha+1)(\alpha-1)!},$$

where the j sum may be calculated using the method of differences. Thus the expression (11) for the expected number of cards turned over can be written

$$T_q=52\sum_{p=1}^{q-1}P(13,p)+212\sum_{\alpha=1}^{13}\frac{f(\alpha-1,q-1)}{(4\alpha+1)(\alpha-1)!}. \tag{13}$$

The values of T_q are tabulated below:

q	T_q
1	42.40000
2	48.71072
3	50.90271
4	51.68720
5	51.92859
6	51.98742
7	51.99834
8	51.99984
9	51.99999
10	52.00000
11	52.00000
12	52.00000
13	52.00000

This shows that the expected number of cards turned converges fairly rapidly to 52 as the number of plays increases.

The authors wish to thank the editor and referees for their suggestions.

Reference

[1] David Kleiner, The last 1 (solution to Problem 1066), this MAGAZINE, 53 (1980) 184–185.

The Evidence for Fortune's Conjecture

SOLOMON W. GOLOMB
University of Southern California
Los Angeles, CA 90007

Euclid's proof that there are infinitely many primes is based on the observation that if $E_n = P_n + 1$ is not itself prime (where $P_n = p_1 p_2 \cdots p_n$ is the product of the first n primes), it must still contain a prime factor larger than p_n.

Little is known about the values of n for which E_n is prime. In fact, the only known prime values of E_n occur for $n = 1, 2, 3, 4, 5$, and 11. When asked by a student whether E_n is prime for infinitely many values of n, George Pólya is reported to have replied, "There are many questions which fools can ask that wise men cannot answer."

The anthropologist Reo Fortune (once married to Margaret Mead) conjectured that if Q_n is the smallest prime number strictly greater than E_n, then the difference $F_n = Q_n - P_n$ is *always* prime. (This conjecture first appeared in print [**2**] in 1980, and is discussed further in [**3**].)

To illustrate, $E_7 = (2 \times 3 \times 5 \times \cdots \times 17) + 1 = 510511$, and the next larger prime is $Q_7 = 510529$. Sure enough, the difference is $F_7 = Q_7 - P_7 = 510529 - 510510 = 19$, a prime. The sequence $\{F_n\}$ of "fortunate numbers" begins: $3, 5, 7, 13, 23, 17, 19, 23, 37, 61, 67, 61, 71, 47, 107, 59, 61, 109, 89, 103, 79, \ldots$, and indeed, all the listed numbers are prime. Is this merely a remarkable coincidence?

At first glance, most mathematicians are tempted to dismiss this conjecture as "almost certainly false." However, a closer inspection reveals that it is quite likely to be true. Since Q_n is known to be prime, $F_n = Q_n - P_n$ cannot be divisible by $p_1, p_2, \ldots, p_n$. Thus $F_n \geqslant p_{n+1}$ for all n, with equality observed at $n = 1,2,3,6,7,8,14,16,17,\ldots$. On the other hand, so long as $F_n < p_{n+1}^2$, F_n *must* be prime, since the smallest composite number coprime to P_n is p_{n+1}^2. It is quite unlikely that F_n is ever as large as p_{n+1}^2, as we shall now see.

Each of the following facts is a consequence of the Prime Number Theorem:

(1) $p_n \sim n \ln n$ as $n \to \infty$.

(2) $\ln P_n = \sum_{i=1}^{n} \ln p_i \sim n$ as $n \to \infty$.

(3) The "expected value" of $p_{n+1} - p_n$ is approximately $\ln n$.

Thus, if P_n were a randomly chosen number, which it is not, the next prime larger than $P_n + 1$ would be expected around $P_n + \ln P_n \sim P_n + n$. However, we have also shown that this next prime Q_n is at least

$$P_n + p_{n+1} \sim P_n + n \ln n.$$

Also, p_{n+1} is at least $p_n + 2$, so that

$$p_{n+1}^2 \geqslant p_n^2 + 4p_n + 4,$$

from which

$$P_n + p_{n+1}^2 > P_n + p_n^2 \sim P_n + n^2 \ln^2 n.$$

Since the *average* separation between primes in the vicinity of P_n is $\ln P_n \sim n$, we *expect* about $n \ln^2 n$ primes in the interval between $P_n + 1$ and $P_n + p_{n+1}^2$. Only if this interval is devoid of primes is it possible for the Fortune conjecture to fail. Indeed, for *every* prime q in the interval between $P_n + 1$ and $P_n + p_{n+1}^2$, we will have $q - P_n$ a prime. TABLE 1 shows the first five values of P_n, the primes q on the interval $(P_n + 1, P_n + p_{n+1}^2)$, and the (necessarily prime) differences $q - P_n$, the first of which is Fortune's number F_n.

n	P_n	$q - P_n = r$ for $q \in (P_n + 1, P_n + p_{n+1}^2)$.	First prime $q^* > P_n$ such that $q^* - P_n$ is composite.
1	2	$5-2=3, 7-2=5$	$11-2=9=p_2^2$
2	6	$11-6=5$, $13-6=7, 17-6=11, 19-6=13, 23-6=17,$ $29-6=23$	$31-6=25=p_3^2$
3	30	$37-30=7$, $41-30=11$, $43-30=13$, $47-30=17$, $53-30=23$, $59-30=29, 61-30=31$, $67-30=37$, $71-30=41$, $73-30=43$	$79-30=49=p_4^2$
4	210	$223-210=13$, $227-210=17$, $229-210=19$, $233-210=23$, $239-210=29$, $241-210=31$, $251-210=41$, $257-210=47$, $263-210=53$, $269-210=59$, $271-210=61$, $277-210=67$, $281-210=71$, $283-210=73$, $293-210=83$, $307-210=97$, $311-210=101, 313-210=103, 317-210=107$.	$331-210=121=p_5^2$
5	2310	$2333-2310=23$, $2339-2310=29$, $2341-2310=31$, $2347-2310=37$, $2351-2310=41$, $2357-2310=47$, $2371-2310=61$, $2377-2310=67$, $2381-2310=71$, $2383-2310=73$, $2389-2310=79$, $2393-2310=83$, $2399-2310=89$, $2411-2310=101$, $2417-2310=107$, $2423-2310=113$, $2437-2310=127$, $2441-2310=131$, $2447-2310=137$, $2459-2310=149$, $2467-2310=157$, $2473-2310=163$, $2477-2310=167$, $\{2503-2310=193$, $2521-2310=211\}$.	$2531-2310=221=p_6p_7$ (The brackets indicate that $169=p_6^2$ is skipped over.)

TABLE 1. **The differences $q - P_n$ for primes $q > P_n + 1$.**

The fact that we *expect* many primes on the interval $(P_n + 1, P_n + p_{n+1}^2)$ is, however, no guarantee that there will always be at least one. A result of the type that there is always a prime on the interval $(x, x + \frac{1}{2}(\ln x \cdot \ln\ln x)^2)$, for all sufficiently large x, is far beyond what can currently be proved, and is quite possibly too strong to be true. Of course, Fortune's Conjecture holds provided that this is true for the very special case when $x = P_n + 1$, for $n = 1,2,3,4,\ldots$.

A well-known conjecture of Cramér [**1**] postulates that the gap between primes in the vicinity of a point x never exceeds $(\ln x)^2$. Thus Cramér's Conjecture implies Fortune's, at least for large n. (To be safe, we can interpret "large n" to mean the region where $\frac{1}{2}\ln\ln n > 1$, which corresponds to $n > [e^{e^2}] = 1618$. Fortune's Conjecture has been verified well beyond $n = [e^e] = 15$, but not yet as far as $n = [e^{e^2}]$.) Of course, no proof of Cramér's Conjecture is currently in sight, though the computer evidence obtained to date is favorable.

Even if there exists an interval $(P_n + 1, P_n + p_{n+1}^2)$ which is free of primes, this does not automatically mean that Fortune's Conjecture will fail for that value of n, because $P_n + p_{n+1}^2$ may itself be composite, as for example $P_5 + p_6^2 = 2310 + 169 = 2479 = 37 \cdot 67$, and this hazard may be passed safely. Thus it is a somewhat stronger conjecture than Fortune's to suppose that the interval strictly between $P_n + 1$ and $P_n + p_{n+1}^2$ always contains at least one prime.

Since $P_8 < 10^7$, D. N. Lehmer's table of primes to ten million [4] can be used to extend TABLE 1 up to $n = 8$. As for *proving* Fortune's Conjecture, however, the reader is cautioned that it is the consensus of number theorists who have considered the problem, that a proof is beyond the reach of current techniques.

This research was supported in part by the United States Army Research Office under Grant No. DAAG 29-79-C-0054.

References

[1] Harald Cramér, On the order of magnitude of the difference between consecutive prime numbers, Acta Arith., 2(1937) 23–46.

[2] Martin Gardner, Mathematical games, Sci. Amer. (December 1980) 18–28. (Column based on contributions by R. K. Guy.)

[3] Richard K. Guy, Unsolved Problems in Intuitive Mathematics, Vol. I, Number Theory, Springer, 1981.

[4] Derrick N. Lehmer, List of Prime Numbers from One to 10,006,721, Publication No. 165, Carnegie Institute of Washington, 1914.

PROBLEMS

Dan Eustice, Editor
Leroy F. Meyers, Associate Editor
The Ohio State University

Proposals

To be considered for publication, solutions should be mailed before March 1, 1982.

1127. (a) Color a 7×7 board with seven colors so that there are exactly seven squares of each color, with no more than three colors in each row and column.

(b)* Color an $n \times n$ board with n colors so that there are exactly n squares of each color, with fewer than $\sqrt{n}+1$ colors in each row and column. [*J. L. Selfridge, Mathematical Reviews.*]

1128. Show that for each $\varepsilon > 0$ there exists a one-to-one mapping σ from the open disk $D = \{z: |z| < 1\}$ onto its closure $\bar{D}$ such that $|z - \sigma(z)| < \varepsilon$ for all z in D. [*Carl P. McCarty & Loretta McCarty, LaSalle College.*]

1129. Find radii r and R, with $r < R$, so that m circles of radius r form a closed ring with each circle externally tangent to a circle of radius R, and n circles of radius r form a closed ring with each circle internally tangent to the circle of radius R. [*Anon, Erewhon-upon-Spanish River.*]

1130. A positive integer is *abundant* if the sum of its proper divisors exceeds n (i.e., $\sigma(n) > 2n$). Show that every integer greater than 89×315 is the sum of two abundant numbers. [*J. L. Selfridge, Mathematical Reviews.*]

1131. Every odd prime is of the form $p = 4n \pm 1$.
(a) Show that n is a quadratic residue (mod p).
(b) Calculate the value of n^n(mod p). [*Oren N. Dalton, El Paso, Texas.*]

ASSISTANT EDITORS: Don Bonar, *Denison University;* William A. McWorter, Jr., *The Ohio State University. We invite readers to submit problems believed to be new. Proposals should be accompanied by solutions, when available, and by any information that will assist the editors. Solutions to published problems should be submitted on separate, signed sheets. An asterisk (*) will be placed by a problem to indicate that the proposer did not supply a solution. A problem submitted as a Quickie should be one that has an unexpected succinct solution. Readers desiring acknowledgment of their communications should include a self-addressed stamped card. Send all communications to this department to Dan Eustice, The Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210.*

Solutions

Real but not Complex **March 1980**

1093. Prove that for complex numbers u, v, and w,

$$|u-v|+|u+v-2w+|u-v||<|u+v| \tag{1}$$

if and only if

$$|w-v|+|w+v-2u+|w-v||<|w+v|. \tag{2}$$

[*M. S. Klamkin, University of Alberta.*]

Editor's Comment. While the result is true if the numbers are real, it is not necessarily true if the numbers are complex. J. M. Stark provided an example which can be simplified to $u=3i$, $v=2i$, and $w=i$, for which the first inequality is true, but the second is false.

Time After Time **May 1980**

1096. An electric timer runs continuously when the current is on, but the outlet from the timer transfers that current (the timer is "on") only during a preset time interval in a 24-hour day. Suppose n electric timers are connected in series (the kth timer is plugged into the outlet of the $(k-1)$st timer) and the kth timer is set to be on for a random time interval of length L_k. If an electric light is plugged into the last timer in the series, what is the expected value for the length of time the light will be on during a 24-hour day? [*Peter Ørno, The Ohio State University.*]

Solution: Let E_k be the expected number of hours per day that the kth timer is on (that is, passing current). Then the solution is the value of E_n.

Let T_k represent the series of the first k timers. Then the probability P_k that T_k is passing current at any moment is $E_k/24$. But

$$P_k = P(k\text{th timer is on} \mid T_{k-1} \text{ is on}) \cdot P_{k-1} = \frac{L_k}{24} P_{k-1}.$$

We thus have a recursive formula for calculating P_n. Noting that $P_0 = 1$ (the probability that the wall outlet is on), we find that

$$P_n = 24^{-n} \prod_{i=1}^{n} L_i,$$

so

$$E_n = 24 P_n = 24^{1-n} \prod_{i=1}^{n} L_i.$$

Chico Problem Group
California State University at Chico

Also solved by Anon (Erewhon), Paul F. Cohen, Bruce R. Johnson, and the proposer. John S. de Cani & Abba M. Krieger and Milton P. Eisner viewed the L_k's as random variables. With this interpretation the expected value is $24/2^n$ hours.

B a Winner **May 1980**

1098. Player A flips $n+1$ coins and keeps n of the coins to maximize the number of heads. Player B flips n coins. The maximum number of heads wins, with ties awarded to Player B. Which player should win and what is the probability of winning? [*Peter Ørno, The Ohio State University.*]

Solution: The game is equivalent to the following: both A and B flip n coins. If either player has more heads than the other, then that player wins. At this point they each have an equal chance of winning. If they both have the same number of heads, but not all heads, A flips his $(n+1)$st coin, winning if it is a head and losing if it is a tail. At this point they still have equal chances of winning. In the remaining case both A and B have all heads. In this case B wins no matter what A's $(n+1)$st flip is. Thus B wins in exactly two more cases than A (all heads or all heads except for A's $(n+1)$st flip). Since there are 2^{2n+1} total flips, the probability that B wins is

$$\frac{2+(2^{2n+1}-2)/2}{2^{2n+1}}=\frac{2^{2n}+1}{2^{2n+1}}.$$

HARRY SEDINGER
St. Bonaventure University

Also solved by Anon (Erewhon), Walter Bluger (Canada), Richard Burns, John S. de Cani & Abba M. Krieger, Chico Problem Group, Jordi Dou (Spain), Philip M. Dunson, Glenn Forney, Nick Franceschine, W. W. Funkenbusch, Heather A. Gamber, Michael Goldberg, Clifford H. Gordon, Richard A. Groeneveld, Joel K. Haack, Victor Hernandez (Spain), G. A. Heuer, Keith Hodge, Bruce R. Johnson (Canada), Krishnamoorthy, John E. Morrill, Roger B. Nelsen, F. D. Parker, Delmar E. Searls, Roy St. Laurent, St. Olaf Problem Solving Group, Philip D. Straffin, Jr., Michael Vowe (Switzerland), Harald Ziehms (Federal Republic of Germany), and Paul Zwier.

AA^* and A^*A **May 1980**

1099. Let A be an $n\times n$ complex matrix and $f(t)$ a polynomial with complex coefficients such that $f(AA^*)=0$. Give an elementary proof that $f(A^*A)=0$. [*Anon, Erewhon-upon-Spanish River.*]

Solution: I: Set $B=f(AA^*)$ and $C=f(A^*A)$. Also, let $g(t)=f(t)\overline{f(t)}$. Then $BB^*=g(AA^*)$ and $CC^*=g(A^*A)$. Hence, because the trace is linear and $\mathrm{tr}(AA^*)^k=\mathrm{tr}(A^*A)^k$, we have $\mathrm{tr}(BB^*)=\mathrm{tr}[g(AA^*)]=\mathrm{tr}(CC^*)$. Since $B=0$, we have $\mathrm{tr}(CC^*)=0$. This implies $C=0$.

ANON
Erewhon-upon-Spanish River

Solution: II: We can use the polar decomposition to write $A=HU$, with U unitary and H Hermitian. Then, $A^*A=U^*H^2U=U^*(AA^*)U$, and so $f(AA^*)=U^*f(AA^*)U=0$.

CHICO PROBLEM GROUP
California State University at Chico

Editor's Comment. Since the polar decomposition is not an elementary result, Solution II, while being short, probably cannot be considered elementary.

Also solved by C. S. K. Chetty (India), Mark Kruelle, and Hans Melissen (The Netherlands).

1100. Suppose that $F(x)$ is a power series (finite or infinite) with rational coefficients and $A_k = \int_0^1 x^k F(x)\,dx$ for integers $k \geqslant 0$.

(i) If all the A_k's are rational, must $F(x)$ be a polynomial?
(ii) Does there exist an $F(x)$ such that all the A_k's except one are rational?
(iii) Does there exist an $F(x)$ such that all the A_k's except $A_{P(k)}, k = 0,1,2,\ldots,$ are rational, where $P(k)$ is an integer valued polynomial, e.g., $P(k) = 2k$?
(iv)*Given a finite indexed set $A_{m(k)}, k = 0,1,2,\ldots, n$, does there exist an $F(x)$ such that all the A_k's except the $A_{m(k)}$'s are rational? [*M. S. Klamkin & M. V. Subbarao, University of Alberta.*]

Solution: Note first that the series $\sum_{i=0}^{\infty} 1/(k+i+1)(m+i+1)$, where k and m are nonnegative integers, is rational if $k \neq m$ and irrational if $k = m$. This follows from the fact that if $k \neq m$,

$$\frac{1}{(k+i+1)(m+i+1)} = \frac{1}{m-k}\left[\frac{1}{k+i+1} - \frac{1}{m+i+1}\right]$$

so that the series is telescopic and, if $k = m$, converges to an irrational number since the series

$$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}.$$

We can now answer questions (i), (ii), and (iv). If we define $c_0 = 0$ and $c_i = 1/i$ then, with $F(x) = \Sigma c_i x^i$,

$$A_k = \sum_{i=1}^{\infty} \frac{1}{i(k+i+1)}$$

which is rational. This answers (i) in the negative. If we define $c_i = 1/(m+i+1)$, and then $F_m(x) = \Sigma c_i x^i$,

$$A_k = \sum_{i=0}^{\infty} \frac{1}{(k+i+1)(m+i+1)}$$

which is rational for $k \neq m$ and irrational for $k = m$. This gives a positive answer to (ii). If we are given a finite set $\{m_1, m_2, \ldots, m_n\}$, then $F_{m_1} + F_{m_2} + \cdots + F_{m_n}$ is a power series with the corresponding A_k's rational except when k is one of the m_i's. This answers (iv) in the affirmative.

As to part (iii), if we take $F(x) = (1-x^2)^{-1/2}$, then F has a power series expansion with rational coefficients and

$$A_k = \int_0^1 x^k (1-x^2)^{-1/2}\,dx.$$

We find $A_0 = \pi/2, A_1 = 1$, and for $k \geqslant 2$,

$$A_k = \frac{k-1}{k} A_{k-2}.$$

Thus A_{2k} is irrational and A_{2k+1} is rational.

PAUL J. ZWIER
Calvin College

Also solved by Chico Problem Group. Parts (i), (ii), and (iii) were solved by the proposers and part (i) by L. Kuipers (Switzerland).

REVIEWS

PAUL J. CAMPBELL, Editor
Beloit College

PIERRE J. MALRAISON, JR., Editor
MDSI, Ann Arbor

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

Axelrod, Robert and Hamilton, William D., *The evolution of cooperation*, Science 211 (27 March 1981) 1390-1396.

Investigates the Prisoner's Dilemma game in the context of evolutionarily stable strategies, and shows that the tit-for-tat strategy is robust, stable, and can gain a foothold.

Sutton, C., *Forests and numbers and thinking backwards*, New Scientist 90 (23 April 1981) 209-212.

A review of the work of J.C.P. Miller on trees, tessellations and number theory.

How the zebra got its spots, New Scientist 90 (28 May 1981) 506-507.

A mathematical theory of how coloration develops in animals. Zebras have stripes rather than spots because the color is determined at an early stage in the development of the fetus.

Bland, R.G., *Allocation of resources by linear programming*, Scientific American 244:6 (June 1981) 126-145, 202.

A look at linear programming, the simplex method, the newly developed and much ballyhooed ellipsoid method, and a discussion of the types of problems amenable to linear programming solutions. The simplex method is theoretically unsatisfying: there exist problems for which it is much slower than the ellipsoid methods. However, for real world problems it still seems to perform better.

Gardner, M., *Mathematical Games*, Scientific American 244:6 (June 1981) 22-32, 202.

A review of Inversions, a book, and other work of Scott Kim. Inversions deals with symmetries in calligraphy invented by Kim. The other work discussed is a generalization of the n queens problem and a question about filling space with polytopes called snakes.

Pattis, Richard E., Karel the Robot: A Gentle Introduction to the Art of Programming, Wiley, 1981; xi + 106 pp + loose errata sheet, $5.95(P).

Presents a short but accurate overview of the terrain of computer programming, in the context of controlling movements of a "robot" on the CRT screen. Excellent LOGO-like one or two week introduction to programming, suitable for the start of a programming course. Karel's language structures are modelled on Pascal; a Karel simulator program (in Pascal) is available for $750 for use on some mini-computers (but not on microcomputers).

Bernstein, H.J. and Philips, A.V., *Fiber bundles and quantum theory*, Scientific American 245:1 (July 1981) 122-137, 188.

A beautiful exposition of how fiber bundles fit into the models used by quantum mechanics. For example, a moebius band is the total space for a fiber bundle over the circle. To get back to where you started when traveling around a moebius band, you have to go 720° around the circle rather than just 360°. A similar model explains the need to rotate neutrons 720° to recover their initial spin. A more complicated model is needed to discuss the Aharonov-Bohm experiment--using a 3-dimensional base space and 4-dimensional total space, along with the notion of a connection (a generalization of curvature) the authors provide another example of the "unreasonable effectiveness" of mathematics.

Hofstadter, D.R., *Metamagical Themas: Pitfalls of the uncertainty principle and paradoxes of quantum mechanics*, Scientific American 245:1 (July 1981) 18-31, 188.

What the uncertainty principle is (and is not) and the paradox of Schrodinger's cat. A nice complement to the article reviewed above.

Grabiner, Judith V., The Origins of Cauchy's Rigorous Calculus, MIT Pr, 1981; 252 pp, $25.

In the eighteenth century analysts concentrated on developing the uses of the calculus; in the nineteenth century the focus was on its justification, as a set of logically derived theorems based on well-defined concepts. The most important figure in the latter development was Cauchy, and this book carefully sets out both his achievements and the influences on him.

Kennedy, Hubert C., Peano: Life and Works of Guiseppe Peano, Reidel, 1980; xii + 231 pp, $34, $14.95(P).

First full-scale study of the life and works of Peano, featuring his two main contributions: his development of mathematical logic (including the Peano postulates) and his efforts to create an international language (Interlingua). Also discussed are the Peano curve, the existence theorem for differential equations, the first axiomatic definition of a vector space, and a new definition of measure. Details of Peano's personal life and sympathies are included.

Amma, T.A. Sarasvati, Geometry in Ancient and Medieval India, Motilal Banarsidass, 1979; xi + 280 pp, $18.50.

Complements the study of arithmetic and algebra in Datta and Singh's *History of Hindu Mathematics* by dealing in detail with Indian geometry from the Vedic Sulbasutras to the 17th century.

Dick, Auguste, Emmy Noether 1882-1935, Birkhäuser, 1981; xiv + 193 pp, $12.95.

Revised translation of Dick's splendid biography of Noether, in an elegant edition. Includes obituaries by van der Waerden, Weyl and Alexander, as well as a list of Noether's publications, a list of the doctoral dissertations she conducted, and a number of photographs.

Stewart, Ian and Jaworski, John (eds.), Seven Years of *Manifold* 1968-1980, Shiva, 1981; ii + 94 pp, £5.

Selections from the most entertaining and whimsical journal of the "undergraduate mathematics society" genre. It aimed to make mathematics accessible, interspersing serious articles with songs, poems and cartoons. The same editors stayed with it to its "Farewell Issue #20," by which time they "had all built careers for themselves and had less time for the buccaneering spirit that had built *Manifold*;" but they seem unwilling to let go, as the order form for the book solicits contributions for future issues! Especially if you never heard of *Manifold* before, don't miss this distillation of it, including "15 new ways to catch a Lion," "Sanctification and the Hopf Bifurcation," "Lemmawocky," "The *Manifold* guide to handwaving," and "Knit Yourself a Klein Bottle."

Bridgman, George, Lake Wobegon Math Problems, available from the author (4306 Grimes Ave. South, Minneapolis, MN 55424); v + 47 pp.

"What are 'Lake Wobegon Math Problems?' They are standard mathematics problems featuring places and people from [the mythical] Lake Wobegon, Minnesota, The Little Town Time Forgot and the Decades Cannot Improve." (The Lake Wobegon motif is taken from Garrison Keillor's Minnesota Public Radio show *A Prairie Home Companion*.) The problems involve high school mathematics and calculus, and the zany dressing may delight students, enhance their motivation, and--if they're not from the Northland--give them wild ideas of what life's like there!

McLeod, Robert M., The Generalized Riemann Integral, Carus Mathematical Monograph 20, MAA, 1980; xiii + 275 pp, $18.

"Rather recently it has been found that an innocent-looking change in the limit process used in the Riemann integral yields an integral with the range and power of the Lebesque integral.... !T1he generalized Riemann integral...can be the first integral defined in calculus courses. It can also be the powerful integral which makes light work of problems requiring interchange of integrals and limit operations...."

Packel, Edward, The Mathematics of Games and Gambling, New Mathematical Library 28, MAA, 1981; x + 137 pp, $8.75(P).

Introduces and develops handily the mathematics needed for elementary analysis of various gambling and social games. Some game theory is included, and the only background needed is some high school algebra. The book grew out of an experimental freshman course over the last three years, and such a course may be an attractive option for many high schools and colleges.

Saunders, P.T., An Introduction to Catastrophe Theory, Cambridge U Pr, 1980; xii + 144 pp.

Excellent mathematical introduction to catastrophe theory for students with a background in multivariable calculus, and a fine component for a course in modelling. The author's attitude toward the controversy over the theory's applications: "...we may claim success not when we have found a catastrophe in Thom's list that matches our observations, but when by doing this (or even, as we have seen, by failing to do this) we have learned something new about the system we are studying."

Gleason, A.M., *et al.*, The William Lowell Putnam Mathematical Competition Problems and Solutions: 1938-1964, MAA, 1980; xi + 652 pp, $35.

In addition to the problems and their solutions for the first 25 contests, the volume contains four articles from the *Monthly* dealing with the founding of the competition, its results, and analysis of its purpose and strengths.

Basic Library List for Two-Year Colleges, 2nd ed., MAA, 1980; 66 pp, $9(P).

Update of the 1971 edition. Four-year colleges will also want to take note, as many of the abooks listed here have appeared in the five years since the last edition of the four-year college list.

Carpenter, Thomas P., *et al.*, Results from the Second Mathematics Assessment of the National Assessment of Educational Progress, National Council of Teachers of Mathematics, 1981; v + 167 pp, $12.50(P).

The second assessment, following the first in 1973, was completed in 1977-78. Performance remained unchanged for 9-year-olds, but declined for 13- and 17-year olds. Highlights: high level of mastery of computational skills, lack of understanding of basic concepts and difficulty with any non-routine problem that required analysis or thinking. "It appears that students have not learned basic problem-solving skills;" in light of which the emphasis on problem-solving of NCTM's "Agenda for Action" for the 80's is right on target. Appendices treat minorities and sex-related differences in mathematics.

NEWS & LETTERS

ALLENDOERFER, FORD, PÓLYA AWARDS

Authors of nine expository articles published in MAA journals in 1980 received awards at the 1981 August meeting of the Association at the University of Pittsburgh. Each award is in the amount of $100.

The recipients of the Carl B. Allendoerfer Awards were selected by a committee consisting of Roy Dubisch, Chairman; Edwin F. Beckenbach, *ex-officio*; and Thomas W. Tucker. The recipients for articles published in 1980 were:

Stephen B. Maurer, "The King Chicken Theorems," *Mathematics Magazine* 53 (1980) 67-80.
Donald E. Sanderson, "Advanced Plane Topology From an Elementary Standpoint," *Mathematics Magazine* 53 (1980) 81-89.

The receipients of the Lester R. Ford Awards were selected by a committee consisting of Branko Grünbaum, Chairman; Edwin F. Beckenbach, *ex-officio*; and Peter L. Duren. The recipients for articles published in 1980 were:

Lawrence Zalcman, "Offbeat Integral Geometry," *Monthly* 87(1980) 161-175.
R. Creighton Buck, "Sherlock Holmes in Babylon," *Monthly* 87(1980) 335-345.
B.H. Pourciau, "Modern Multiplier Rules," *Monthly* 87(1980) 433-452.
E.R. Swart, "The Philosophical Implications of the Four-Color Problem," *Monthly* 87(1980) 697-707.
Alan H. Schoenfeld, "Teaching Problem-Solving Skills," *Monthly* 87(1980) 794-805.

The recipients of the George Pólya Awards were selected by a committee consisting of Kay W. Dundas, Chairman; Edwin F. Beckenbach, *ex-officio*; and Warren Page. The recipients for articles published in 1980 were:

E.D. McCune, R.G. Dean, and W.D. Clark, "Calculators to Motivate Infinite Composition of Functions," Vol. 11, No. 3, pp. 189-195, *Two-Year College Mathematics Journal*.
G.D. Chakerian, "Circles and Spheres," Vol. 11, No. 1, pp. 26-41, *Two-Year College Mathematics Journal*.

INVERTING SUMS OF MATRICES

The paper on inverting sums of matrices by Kenneth S. Miller (this *Magazine*, March 1981, pp. 67-72) used a certain matrix identity as a lemma. There is a shorter proof of a more general identity than the one given in the paper. Let G be a nonsingular $n\times n$ matrix and u,v be n-dimensional column vectors. Suppose that $g = v'G^{-1}u$ is not zero, and let $a = g^{-1}$. For any scalar p define the matrix $V(p) = G + a(p-1)uv'$. In particular $V(1) = G$. By multiplying and canceling terms, it is easy to verify that $V(p)\ G^{-1}V(q) = V(pq)$ for any scalars p,q. One consequence of this identity is that any matrix of the form $V(p)$ can be easily inverted if $p \neq 0$: $[V(p)]^{-1} = G^{-1}\ V(p^{-1})\ G^{-1}$. As a special case the matrix $G + uv'$ can be inverted by proper choice of p; this gives the identity in the paper.

While the identity has probably been known for fifty years, the simple proof does not seem to be widely known. The proof shows the essential reason for the power of the identity; namely that the semigroup of matrices of the form $V(p)$, with the operation $A \circ B = AG^{-1}B$, is isomorphic to the semigroup of multiplicative real numbers. That means that one can invert and find square roots of these matrices by just operating on real numbers, which often saves a lot of computing time.

David Morrison
TRW
Redondo Beach, CA 90278

NESTED OSCULATING CIRCLES

Theorem 1 in the paper "Nesting Behavior of Osculating Circles..." by Joel Zeitlin (this *Magazine*, March 1981, pp. 76-78) states, "If $\alpha(s)$ is a regular curve with radius of curvature $\rho(s)$ and $\rho(s) \neq 0$ for $s \in I$ (some interval), then no two of the osculating circles at $\alpha(s)$

for $s \in I$ intersect, *i.e.*, the circles are nested within one another."

In the paper, this theorem is attributed to Stoker, 1969. Earlier, the theorem was stated and proved in A.A. Blank, Problems in Calculus and Analysis, Wiley, NY, 1966, as part of the solution of Problem 9 Section 4.1h (*op. cit.*, pp. 186, 189). If anyone knows an earlier reference, I shall appreciate being informed of it.

Albert A. Blank
Carnegie-Mellon University
Pittsburgh, PA 15213

AUTHOR'S REPLY

A version of Theorem 1 in my *Mathematics Magazine* paper appears with credit given to Kneser on page 48 of Heinrich Guggenheimer, Differential Geometry, McGraw Hill Book Company, Inc., Copyright 1963. Page 72 gives references for this section 3-3.

Joel Zeitlin
California State University
Northridge, CA 91330

THEOLOGY AND GAMES

In his letter (this *Magazine*, May 1981, p. 148), Ian Richards attacks the assumptions of my analysis (this *Magazine*, November 1980, pp. 227-282)--particularly that God might make strategic choices in a game (including to lie).

I would refer Richards to my book, Biblical Games: A Strategic Analysis of Stories in the Old Testament (MIT Press, 1980), cited in my article, which includes approximately three hundred quotations from the Old Testament in support of the validity of treating God as a game player in biblical exegesis. This work is, I believe, in what Richards calls the "venerable tradition" of discussing problems in the "moral sphere" on the basis of "their historical and philosophical merits." Philosophical analysis also informs a follow-up article, "A Resolution of the Paradox of Omniscience," that I did to my *Mathematics Magazine* article, which will appear in the Proceedings of the Conference on Reason and Decision, Third Annual Conference in Applied Philosophy, Bowling Green State University, May 1-2, 1981.

There is a more general intellectual issue that Richards's letter raises. He claims that assumptions of the kind I make are "false," without saying what truth is. I believe that the "standards of intellectual rigor" that he extols, and which we all strive for in one way or another, are inconsistent with one person's prescribing what truth--or falsity--is.

Steven J. Brams
Department of Politics
New York University
New York, NY 10003

EXHIBIT AT THE FRANKLIN INSTITUTE

A new exhibition entitled "Patterns" will run through January, 1982 at The Franklin Institute Science Museum, **20th** and Parkway, Philadelphia.

The exhibition is divided into four main sections--Symmetry, Perspective, Geometry and Numbers--each providing a different exposure to the relationship between math and the arts.

Visitors will be able to use computers to recreate a painting by Ellsworth Kelley that used random numbers in its composition; create their own attractive symmetry patterns; compose simple music or find out how "serialist" composers used mathematics in their work.

In addition to the many computer programs, "Patterns" offers 12 experimental stations for visitors of all ages.

Music that has a mathematical basis will be presented as well as a resource center stocked with magazine articles, books, tapes with commentary, the mathematics of musical scales using a Mozart composition and much more.

"Patterns" is supported by grants from the National Endowment for the Arts, the National Endowment for the Humanities, the Pennsylvania Council on the Arts and The Franklin Institute Galaxy Ball Committee and several private donors.

XXII INTERNATIONAL MATHEMATICAL OLYMPIAD

WASHINGTON, DC, U.S.A.

On July 13 and 14, 1981, in Washington, DC, teams of high school students from 27 countries participated in the XXII International Mathematical Olympiad. Each country's team had from one to eight students. On each of the two days, participants were given three problems to solve in 4-1/2 hours. The problems are reproduced in the next column.

This year the participants made an exceptionally strong showing on the examination. Twenty-six students produced perfect papers, and these, together with ten others, all received first prize. Countries, together with the number of their students having a perfect paper are as follows: Austria-2; Brazil-1; Bulgaria-1; Canada-2; Czechoslovakia-1; France-2; Federal Republic of Germany-3; Hungary-1; Israel-1; Luxemburg-1; Poland-1; United Kingdom-2; United States-4; USSR-3; Yugoslavia-1.

The three countries with highest team score averages were: first--U.S.A.; second--Federal Republic of Germany; third--United Kingdom.

Students on the U.S. team are listed below. The first four made perfect scores.

Noam D. Elkies
Stuyvesant H.S., New York, NY

Benjamin N. Fisher
Bronx H.S. of Science, New York, NY

Brian R. Hunt
Montgomery Blair H.S., Silver Spring, MD

Jeremy D. Primer
Columbia H.S., Maplewood, NJ

Gregg N. Patruno
Stuyvesant H.S., New York, NY

James R. Roche
Hill-Murray H.S., St. Paul, MN

Richard A. Strong
Albemarle H.S., Charlottesville, VA

David S. Yuen
Lane Technical H.S., Chicago, IL

1. P is a point inside a given triangle ABC. D, E, F are the feet of the perpendiculars from P to the lines BC, CA, AB, respectively. Find all P for which

$$\frac{BC}{PD} + \frac{CA}{PE} + \frac{AB}{PF}$$

is least.

2. Let $1 \leq r \leq n$ and consider all subsets of r elements of the set $\{1, 2, \ldots, n\}$. Also consider the least number in each of these subsets. $F(n, r)$ denotes the arithmetic mean of these least numbers; prove that

$$F(n, r) = \frac{n+1}{r+1}.$$

3. Determine the maximum value of $m^2 + n^2$, where m and n are integers satisfying $m, n \in \{1, 2, \ldots, 1981\}$ and $(n^2 - mn - m^2)^2 = 1$.

4.(a) For which values of $n > 2$ is there a set of n consecutive positive integers such that the largest number in the set is a divisor of the least common multiple of the remaining $n - 1$ numbers?

(b) For which values of $n > 2$ is there exactly one set having the stated property?

5. Three congruent circles have a common point O and lie inside a given triangle. Each circle touches a pair of sides of the triangle. Prove that the incenter and the circumcenter of the triangle, and the point O are collinear.

6. The function $f(x, y)$ satisfies

(1) $f(0, y) = y+1$,

(2) $f(x+1, 0) = f(x, 1)$,

(3) $f(x+1, y+1) = f(x, f(x+1, y))$,

for all nonnegative integers x, y. Determine $f(4, 1981)$.

Eminent Mathematicians and Mathematical Expositors speak to

STUDENTS and TEACHERS in . . .

The NEW MATHEMATICAL LIBRARY

An internationally acclaimed paperback series providing:

- stimulating excursions for students beyond traditional school mathematics
- supplementary reading for school and college classrooms
- valuable background reading for teachers
- challenging problems for solvers of all ages from high school competitions in the US and abroad

The New Mathematical Library is published by the MATHEMATICAL ASSOCIATION OF AMERICA. The volumes are paperbound.

For information regarding the price of these publications, please contact The Mathematical Association of America at the address listed below.

NUMBERS: RATIONAL AND IRRATIONAL by Ivan Niven, NML-01

WHAT IS CALCULUS ABOUT? By W. W. Sawyer, NML-02

AN INTRODUCTION TO INEQUALITIES, by E. F. Beckenbach, and R. Bellman, NML-03

GEOMETRIC INEQUALITIES, By N. D. Kazarinoff, NML-04

THE CONTEST PROBLEM BOOK. Problems from the Annual High School Mathematics Contests sponsored by the MAA, NCTM, Mu Alpha Theta, The Society of Actuaries, and the Casualty Actuarial Society. Covers the period 1950-1960. Compiled and with solutions by C. T. Salkind. NML-05

THE LORE OF LARGE NUMBERS, by P. J. Davis, NML-06

USES OF INFINITY, by Leo Zippin, NML-07

GEOMETRIC TRANSFORMATIONS, by I. M. Yaglom, translated by Allen Shields, NML-08

CONTINUED FRACTIONS, by C. D. Olds, NML-09

GRAPHS AND THEIR USES, by Oystein Ore, NML-10

HUNGARIAN PROBLEM BOOKS I and II, based on the Eötvos Competitions 1894-1905 and 1906-1928. Translated by E. Rapaport, NML-11 and NML-12

EPISODES FROM THE EARLY HISTORY OF MATHEMATICS, by A. Aaboe, NML-13

GROUPS AND THEIR GRAPHS, by I. Grossman and W. Magnus, NML-14

THE MATHEMATICS OF CHOICE, by Ivan Niven, NML-15

FROM PYTHAGORAS TO EINSTEIN, by K. O. Friedrichs, NML-16

THE CONTEST PROBLEM BOOK II. A continuation of NML-05 containing problems and solutions from the Annual High School Mathematics Contests for the period 1961-1965. NML-17

FIRST CONCEPTS OF TOPOLOGY, by W. G. Chinn and N. E. Steenrod, NML-18

GEOMETRY REVISITED, by H.S.M. Coxeter, and S. L. Greitzer, NML-19

INVITATION TO NUMBER THEORY, by Oystein Ore, NML-20

GEOMETRIC TRANSFORMATIONS II, by I. M. Yaglom, translated by Allen Shields, NML-21

ELEMENTARY CRYPTANALYSIS — A Mathematical Approach, by Abraham Sinkov, NML-22

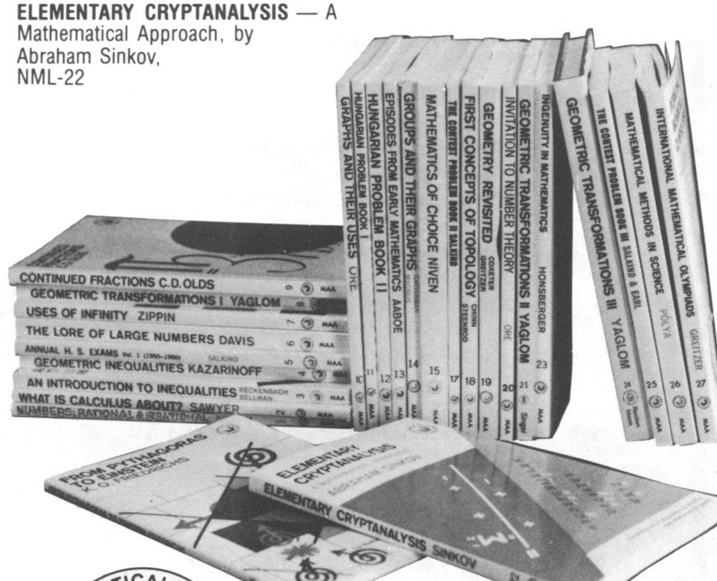
INGENUITY IN MATHEMATICS, by Ross Honsberger, NML-23

GEOMETRIC TRANSFORMATIONS III, by I. M. Yaglom, translated by Abe Shenitzer, NML-24

THE CONTEST PROBLEM BOOK III. A continuation of NML-05 and NML-17, containing problems and solutions from the Annual High School Mathematics Contests for the period 1966-1972. NML-25

MATHEMATICAL METHODS IN SCIENCE, by George Pólya, NML-26

INTERNATIONAL MATHEMATICAL OLYMPIADS, 1959-1977. Problems, with solutions, from the first nineteen International Mathematical Olympiads. Compiled and with solutions by S. L. Greitzer. NML-27



THE MATHEMATICAL ASSOCIATION OF AMERICA

Send orders to: **The Mathematical Association of America**
1529 Eighteenth St., N.W., Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

MATHEMATICS MAGAZINE VOL. 54, NO. 4, SEPTEMBER 1981