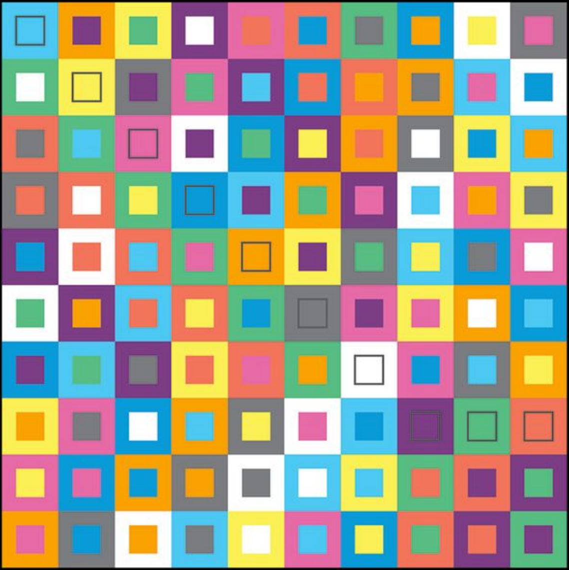
COMBINATORICS:
ANCIENT & MODERN



EDITED BY

ROBIN WILSON & JOHN J. WATKINS

OXFORD

Combinatorics: Ancient and Modern

Title page of Athanasius Kircher's *Ars Magna Sciendi Sive Combinatoria* (The Great Art of Knowledge, or the Combinatorial Art) of 1669.

COMBINATORICS: ANCIENT AND MODERN

Edited by
ROBIN WILSON
and
JOHN J. WATKINS

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published 2013
Impression: 1

British Library Cataloguing in Publication Data
Data available

ISBN 978–0–19–965659–2

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

FOREWORD

One of the most compelling instincts that human beings have is the irresistible urge to look for patterns: this is apparent from the earliest attempts of our ancestors to understand the world around them. Mathematics has often been described as the science of patterns, and perhaps more than any other mathematical field, this represents the heart and soul of combinatorics.

In this masterful volume, the editors have brought together a wonderful and far-ranging collection of chapters by distinguished authors. These accounts survey the subject of combinatorics, beginning with the earliest written results in the subject, continuing with its development in a variety of cultures, such as Indian, Chinese, Islamic, and Jewish, and progressing to the emergence of what we now think of as modern combinatorics. From the introductory chapter, 'Two thousand years of combinatorics', by Donald Knuth to Peter Cameron's 'A personal view of combinatorics', the book covers a wide range of topics and offers to both the novice in the subject and to the experts a full plate of interesting facts and viewpoints. This is the first time that such a compilation has been attempted and, in the opinion of this reader, it succeeds brilliantly.

Ronald Graham
Former President of the American Mathematical Society
and the Mathematical Association of America

This page intentionally left blank

PREFACE

Who first presented Pascal's triangle? (It was not Pascal.)
Who first presented Hamiltonian graphs? (It was not Hamilton.)
Who first presented Steiner triple systems? (It was not Steiner.)
Misattributions appear throughout the history of mathematics, and combinatorics has its share. Frequent errors in historical accuracy are perpetrated by those who would be ashamed to allow such errors in their mathematical writings: notable examples are the common assertions that Euler drew a four-vertex graph to solve the Königsberg bridges problem and that Descartes discovered Euler's polyhedron formula (they didn't).

Today the history of mathematics is a well-studied and vibrant area of research, with books and scholarly articles published on various aspects of the subject. Yet, the history of combinatorics seems to have been largely overlooked: many combinatorialists seem uninterested in the history of their subject, while historians of mathematics have tended to bypass the fascinations of combinatorics. It is our hope that this book, written by noteworthy experts in the area, will go some way to redress this.

This book serves two purposes:

- It constitutes what is perhaps the first book-length survey of the history of combinatorics.
- It assembles, for the first time in a single source, research on the history of combinatorics that would otherwise be inaccessible to the general reader.

The chapters have been contributed by sixteen experts, with topics corresponding to their particular areas of research. Some of this research receives its first airing here, while other chapters are based on work that has appeared elsewhere but is largely unavailable to those without access to research journals or large university libraries. In order to make the book easier to read, we have endeavoured to standardize the style throughout the book, and we are grateful to the authors for their forbearance with us as we proposed changes to their drafts and imposed our stylistic conventions.

The opening section is an introduction to two thousand years of combinatorics, adapted with permission from a section of Volume 4 of Donald E. Knuth's celebrated multi-volume work *The Art of Computer Programming*. This is followed by seven chapters on early combinatorics, leading from Indian and Chinese writings on permutations to late-Renaissance publications on the arithmetical triangle. The next seven chapters trace the subsequent story, from Euler's contributions to such wide-ranging topics as partitions, polyhedra, and latin squares to the 20th-century advances in combinatorial set theory, enumeration, and graph theory. The book concludes with some combinatorial reflections by the distinguished combinatorialist Peter J. Cameron.

Naturally, as the first book of this kind, this volume cannot hope to be comprehensive, and you will notice topics that are missing or only minimally discussed: these include combinatorial optimization, combinatorial identities, and recreational combinatorics. While these subjects (and others) were considered for inclusion we felt that the necessary constraints of a volume of manageable scope and size left too little room. Such omissions notwithstanding, we hope that as an overview of recent and ongoing historical work this book provides useful background information and inspiration for future research.

As with many edited volumes, this book is not intended to be read from cover to cover, although it can be. Rather, it is intended to serve as a valuable resource to a variety of audiences. We hope that combinatorialists with little or no knowledge about the development of their subject will find the historical treatment stimulating, and that the specialist historian of mathematics will view its assorted surveys as an encouragement for further research in combinatorics. For the more general reader, we hope that it provides an introduction to a fascinating and too little known subject that continues to stimulate and inspire the work of scholars today.

Finally, we'd like to express our thanks to Keith Mansfield, Viki Mortimer, and Clare Charles at Oxford University Press, and also to our host institutions, Pembroke College, Oxford University; The Open University, UK; and The Colorado College, USA.

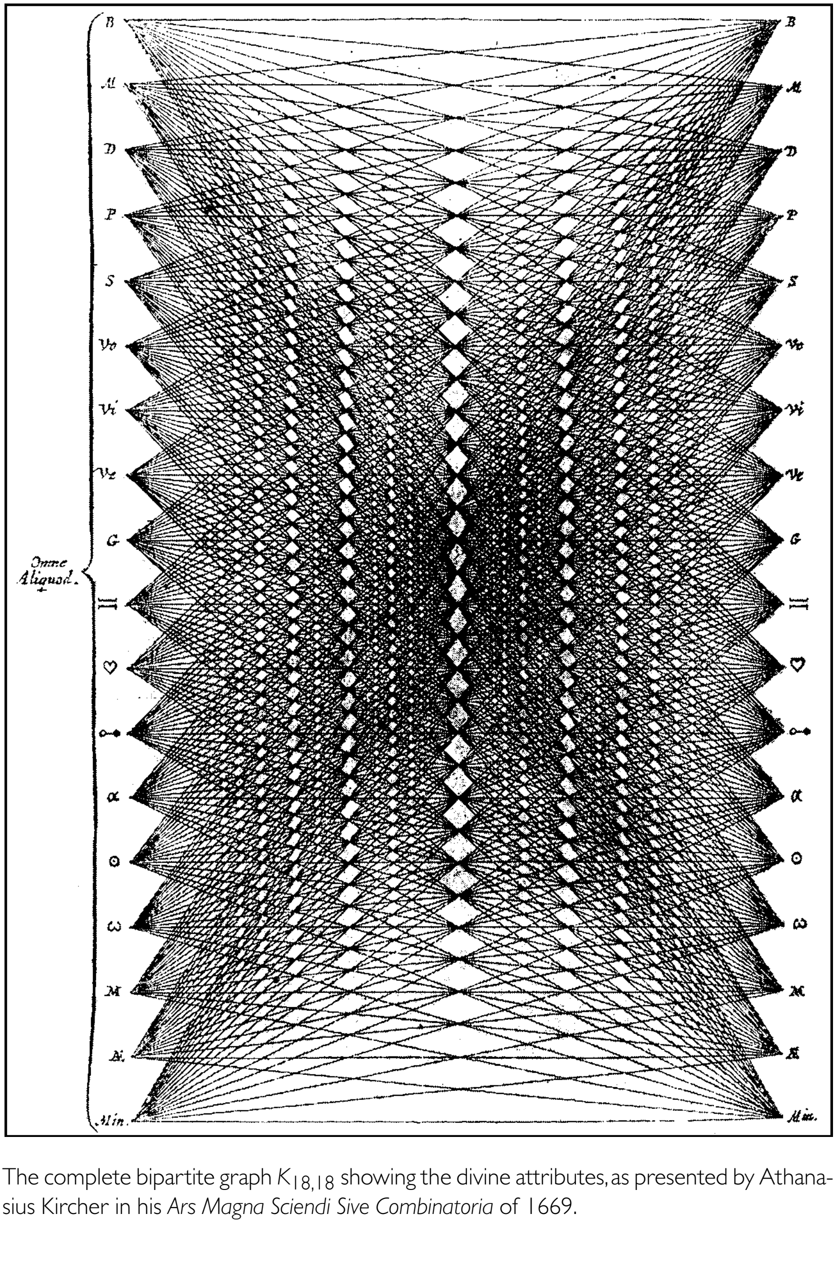
The Editors

April 2013

CONTENTS

PART I

Introduction



The complete bipartite graph $K_{18,18}$ showing the divine attributes, as presented by Athanasius Kircher in his *Ars Magna Sciendi Sive Combinatoria* of 1669.

TWO THOUSAND YEARS OF COMBINATORICS

DONALD E. KNUTH

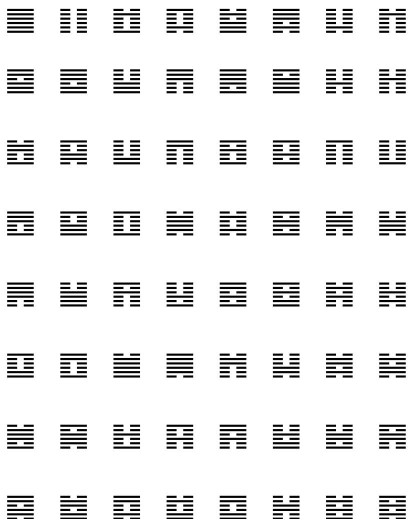
[This subject] has a relation
to almost every species of useful knowledge
that the mind of man can be employed upon.

Jacob Bernoulli, *Ars Conjectandi* (1713)

Early work on the generation of combinatorial patterns began as civilization itself was taking shape. The story is quite fascinating, and we will see that it spans many cultures in many parts of the world, with ties to poetry, music, and religion. There is space here to discuss only some of the principal highlights; but perhaps a few glimpses into the past will stimulate us to dig deeper into the roots of the subject, as the world gets ever smaller and as global scholarship continues to advance.

The *I Ching*

Lists of binary n-tuples can be traced back thousands of years to ancient China, India, and Greece. As we see in Chapter 2, the most notable source – because it still is a best-selling book in modern translations – is the Chinese *I Ching* or *Yijing* (Book of Change). This book, which is one of the five classics of Confucian wisdom, consists essentially of $2^6 = 64$ chapters; and each chapter is symbolized by a hexagram formed from six lines, each of which is either – – (yin) or — (yang). For example, the first hexagram is pure yang ䷀, the second is pure yin ䷁, and the last two hexagrams intermix yin and yang with yin on top in one ䷾, and yang on top in the other ䷿. The standard arrangement of the sixty-four possibilities, shown below, is called the King Wen order, because the basic text of the *I Ching* has traditionally been ascribed to King Wen (*c.*1100 BC), the legendary progenitor of the Chou dynasty. Ancient texts are, however, notoriously difficult to date reliably, and modern historians have found no solid evidence that anyone actually compiled such a list of hexagrams before the 3rd century BC.



Notice that the hexagrams of the King Wen order occur in pairs. Each diagram is immediately followed by its top-to-bottom reflection – for example, ䷂ is followed by ䷃ – except when reflection would make no change; and the eight symmetrical diagrams are paired with their complements (䷀ and ䷁, ䷚ and ䷛, ䷜ and ䷝, ䷼ and ䷽). Hexagrams that are composed from two trigrams, representing the four basic elements of heaven (☰), earth (☷), fire (☲), and water (☵), have also been placed judiciously. Otherwise, the arrangement appears to be essentially random, as if a person untrained in mathematics kept listing different possibilities until being unable to come up with any more. A few intriguing patterns do exist between the pairs, but no more than are present by coincidence in the digits of π.

Yin and yang represent complementary aspects of the elementary forces of nature, always in tension, always changing. The *I Ching* is somewhat analogous to a thesaurus in which the hexagrams serve as an index to accumulated wisdom about fundamental concepts like giving (䷀), receiving (䷁), modesty (䷎), joy (䷹), fellowship (䷌), withdrawal (䷠), peace (䷊), conflict (䷅), organization (䷆), corruption (䷑), immaturity (䷃), elegance (䷕), etc. One can choose a pair of hexagrams at random, obtaining the second from the first by, say, independently changing each yin to yang (or vice versa) with probability $\frac{1}{4}$; this technique yields 4096 ways to ponder existential mysteries, as well as a Markov process by which change itself might perhaps give meaning to life.

A strictly logical way to arrange the hexagrams was eventually introduced around AD 1060 by Shao Yung. His ordering, which proceeded lexicographically from ䷁ to ䷖ to ䷇ to ䷓ to ䷏ to . . . to ䷪ to ䷀ (reading each hexagram from bottom to top), was much more user-friendly than the King Wen order, because a random pattern could now be found quickly. When G. W. Leibniz learned about this sequence of hexagrams in 1702, he jumped to the erroneous conclusion that Chinese mathematicians had once been familiar with binary arithmetic (see [65]). Further details about the *I Ching* can be found in [49] and [41].

Another ancient Chinese philosopher, Yang Hsiung, proposed a system based on eighty-one ternary tetragrams instead of sixty-four binary hexagrams. His *Canon of Supreme Mystery* (around 2 BC) has recently been translated into English by Michael Nylan [72]. Yang described a complete hierarchical ternary tree structure in which there are three regions, with three provinces in each region, three departments in each province, three families in each department, and nine short poems called 'appraisals' for each family, hence 729 appraisals in all – making almost exactly two appraisals for each day in the year. His tetragrams were arranged in strict lexicographic order when read top-to-bottom: 𝌆, 𝌇, 𝌈, 𝌉, 𝌊, 𝌋, 𝌌, . . . , 𝍖. In fact, as explained in [72, p. 28], Yang presented a simple way to compute the rank of each tetragram, as if using a base-3 number system. Thus he would not have been surprised or impressed by Shao Yung's systematic ordering of binary hexagrams, although Shao lived more than a thousand years later.

Indian prosody

Binary n-tuples were studied in a completely different context by pundits in ancient India, who investigated the poetic metres of sacred Vedic chants. As we see in Chapter 1, syllables in Sanskrit are either short or long, and the study of syllable patterns is called ‘prosody’. Modern writers use the symbols ◡ and — to represent, respectively, short and long syllables. A typical Vedic verse consists of four lines with n syllables per line, for some $n \geq 8$; prosodists therefore sought a way to classify all 2^n possibilities.

The classic work *Chandaḥśāstra* by Piṅgala, written before AD 400 and probably much earlier (the exact date is quite uncertain), described procedures by which one could readily find the index k of any given pattern of ◡ s and — s, as well as find the kth pattern, given k. In other words, Piṅgala explained how to *rank* any given pattern as well as to *unrank* any given index; thus he went beyond the work of Yang Hsiung, who had considered ranking but not unranking. Piṅgala’s methods were also related to exponentiation.

The next important step was taken by a prosodist named Kedāra in his work *Vṛttaratnākara*, thought to have been written in the 8th century. Kedāra gave a step-by-step procedure for listing all the n-tuples from ——— . . . — to ◡—— . . . — to —◡— . . . — to ◡◡— . . . — to ——◡ . . . — to ◡—◡ . . . — to $\cdots$ to ◡◡◡ . . . ◡. His method may well have been the first-ever explicit algorithm for combinatorial sequence generation (see [52]).

Poetic metres can also be regarded as rhythms, with one beat for each ◡ and two beats for each —. An n-syllable pattern can involve between n and $2n$ beats, but musical rhythms suitable for marching or dancing are generally based on a fixed number of beats. Therefore it was natural to consider the set of all sequences of ◡ s and — s that have exactly m beats, for fixed m. Such patterns are now called *Morse code sequences* of length m, and it is not at all hard to show that the number of such sequences is the Fibonacci number F_{m+1}; in this way Indian prosodists were led to discover the Fibonacci sequence. For example, the 21 sequences when $m = 7$ are

◡———, —◡——, ◡◡◡——, ——◡—, ◡◡—◡—,
◡—◡◡—, —◡◡◡—, ◡◡◡◡◡—, ———◡,
◡◡——◡, ◡—◡—◡, —◡◡—◡, ◡◡◡◡—◡,
◡——◡◡, —◡—◡◡, ◡◡◡—◡◡, ——◡◡◡,
◡◡—◡◡◡, ◡—◡◡◡◡, —◡◡◡◡◡, ◡◡◡◡◡◡◡.

Moreover, the anonymous author of *Prākṛta Paiṅgala* (*c*.1320) discovered elegant algorithms for ranking and unranking with respect to m-beat rhythms. To find the kth pattern, one starts by writing down m ◡ s, then expresses the difference $d = F_{m+1} - k$ as a sum of Fibonacci numbers $F_{j_1} + \cdots + F_{j_t}$; here F_{j_1} is the largest Fibonacci number that does not exceed d and F_{j_2} is the largest not exceeding $d - F_{j_1}$, etc., continuing until the remainder is zero. Then beats $j-1$ and j are to be changed from ⏑⏑ to —, for $j = j_1, \ldots, j_t$. For example, to get the fifth element of the list above we compute $21 - 5 = 16 = 13 + 3 = F_7 + F_4$; thus, on changing beats 6 and 7, and 3 and 4, the answer is ⏑⏑—⏑—.

A few years later, Nārāyaṇa Paṇḍita treated the more general problem of finding all compositions of m whose parts are at most q, where q is *any* given positive integer. As a consequence he discovered the qth-order Fibonacci sequence, which was destined to be used six hundred years later in polyphase sorting; he also developed the corresponding ranking and unranking algorithms (see [60]).

Piṅgala gave special code names to all the three-syllable metres,

——— ——⏑ ⏑—— ⏑—⏑ —⏑— —⏑⏑ ⏑⏑— ⏑⏑⏑ ,

and students of Sanskrit have been expected to memorize them ever since. Somebody long ago devised a clever way to recall these codes, by inventing the nonsense word *yamātārājabhānasalagām*; the point is that the ten syllables of this word can be written

ya	mā	tā	rā	ja	bhā	na	sa	la	gām
⏑	—	—	—	⏑	—	⏑	⏑	⏑	—

and each three-syllable pattern occurs just after its code name. The origin of this nonsense word is obscure, but Subhash Kak [28] has traced it back at least to C. P. Brown's *Sanskrit Prosody* [12] of 1869; thus it qualifies as the earliest known appearance of a 'de Bruijn cycle' that encodes binary n-tuples. Further results on Indian prosody are given in Chapter 1.

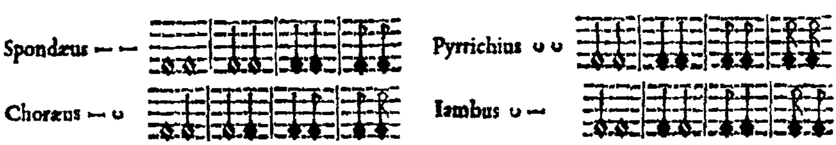
Meanwhile, in Europe

In a similar way, classic Greek poetry was based on groups of short and long syllables called 'metrical feet', analogous to bars of music. Each basic type of foot acquired a Greek name; for example, two short syllables ⏑⏑ were called a *pyrrhic*, and two long syllables —— were called a *spondee*, because those rhythms were used respectively in a song of war ($\pi\upsilon\rho\rho\acute{\iota}\chi\eta$) or a song of peace ($\sigma\pi o\nu\delta\alpha\acute{\iota}$). Greek names for metric feet were soon assimilated into Latin and eventually into modern languages, including English; those names are tabulated below.

◡	arsis	◡◡◡◡	proceleusmatic
—	thesis	◡◡◡—	fourth pæon
		◡◡—◡	third pæon
◡◡	pyrrhic	◡◡——	minor ionic
◡—	iambus	◡—◡◡	second pæon
—◡	trochee	◡—◡—	diiambus
——	spondee	◡——◡	antispast
		◡———	first epitrite
◡◡◡	tribrach	—◡◡◡	first pæon
◡◡—	anapest	—◡◡—	choriambus
◡—◡	amphibrach	—◡—◡	ditrochee
◡——	bacchius	—◡——	second epitrite
—◡◡	dactyl	——◡◡	major ionic
—◡—	amphimacer	——◡—	third epitrite
——◡	palimbacchius	———◡	fourth epitrite
———	molossus	————	dispondee

Alternative terms, such as ‘choree’ instead of ‘trochee’, or ‘cretic’ instead of ‘amphimacer’, were also in common use. Moreover, by the time Diomedes wrote his Latin grammar (around AD 375), each of the thirty-two *five*-syllable feet had acquired at least one name. Diomedes also pointed out the relation between complementary patterns; he stated, for example, that tribrach and molossus are *contrarius*, as are amphibrach and amphimacer. But he also regarded dactyl as the contrary of anapest, and bacchius as the contrary of palimbacchius, although the literal meaning of *palimbacchius* is actually ‘reverse bacchius’ (see [15] and [29]). Greek prosodists had no standard order in which to list the individual possibilities, and the form of the names makes it clear that no connection to a base-2 number system was contemplated.

Surviving fragments of a work by Aristoxenus called *Elements of Rhythm* (*c*.325 BC) show that the same terminology was also applied to music. And indeed, the same traditions lived on after the Renaissance; for example, we find



in Athanasius Kircher's *Musurgia Universalis* [30, p. 32] of 1650, and Kircher went on to describe all of the three-note and four-note rhythms listed above. Kircher's combinatorial writings are studied in greater detail in Chapter 5.

Early lists of permutations

Non-trivial lists of *permutations* were not published until hundreds of years after the formula $n!$ was discovered. The first such tabulation currently known was compiled by the Italian physician Shabbetai Donnolo in his commentary on the kabbalistic *Sefer Yetsirah* (see Chapter 4). The following table shows his list for $n = 5$ as it was subsequently printed in Warsaw in 1884. (The Hebrew letters in this table are typeset in a rabbinical font traditionally used for commentaries.)

דברים, דבירם, דבימר, דבמיר, דבמרי, דברמי, דרבמי, דרבים, דרימב, דריבם, דרמבי, דרמיב,

דיברם, דירבם, דירמב, דיבמר, דימבר, דימרב, דמביר, דמברי, דמרבי, דמריב, דמיבר, דמירב,

בדימר, בדירם, בדרמי, בדרים, בדמרי, בדמיר, בירמד, בידרם, בירדם, בימדר, בימרד, בידמר,

במדרי, במדיר, במריד, במרדי, במדיר, במירד, בדמיר, בדמרי, בדרמי, בדרים, בדימר, בדירם,

רימדב, רימבד, רידבם, רידמב, ריבדם, ריבמד, רדמבי, רדבמי, רדיבם, רדימב, רדבים, רדמיב,

רבמדי, רבדמי, רבדים, רבידם, רבימד, רבמיד, רמידב, רמיבד, רמדיב, רמביד, רמבדי, רמדבי,

ימדרב, ימרדב, ימבדר, ימברד, ימרבד, ימדבר, ידברם, ידרבם, ידבמר, ידמבר, ידרמב, ידמרב,

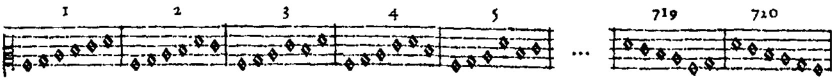
יבדרם, יבדמר, יברדם, יברמד, יבמדר, יבמרד, ירדבם, ירדמב, ירמבד, ירבמד, ירבדם, ירדמב,

מדברי, מדרבי, מדביר, מדיבר, מדירב, מדריב, מבדיר, מבידר, מבריד, מבירד, מברדי, מבדרי,

מרדבי, מרבדי, מרדיב, מרביד, מריבד, מרידב, מידבר, מידרב, מיברד, מיבדר, מירדב, מירבד

A medieval list of permutations.

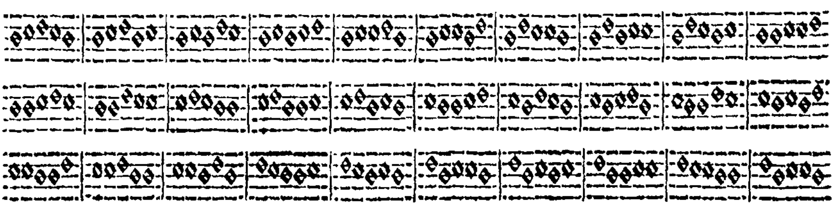
Donnolo went on to list 120 permutations of a six-letter word, all beginning with the letter *shin*; then he noted that 120 more could be obtained with each of the other five letters in front, making 720 in all. His lists involved groupings of six permutations, but in a haphazard fashion that led him into error. Although he knew how many permutations there were supposed to be, and how many should start with a given letter, he evidently had no algorithm for generating them.

A complete list of all 720 permutations of $\{a, b, c, d, e, f\}$ appeared in Jeremias Drexel's *Orbis Phaëthon* in 1629 (see [16] and [17]). Drexel offered it as proof that a man with six guests could seat them differently at lunch and dinner every day for a year – altogether 360 days, because there were five days of fasting during Holy Week. Shortly afterwards, as we see in Chapter 5, Marin Mersenne exhibited all 720 permutations of the six tones {ut, re, mi, fa, sol, la} in his *Traitez de la Voix et des Chants* (Treatise on the Voice and Singing) [44], also presenting the same data in musical notation:

Drexel's table was organized lexicographically by columns; Mersenne's tables were lexicographic with respect to the order ut < re < mi < fa < sol < la, beginning with 'ut, re, mi, fa, sol, la' and ending with 'la, sol, fa, mi, re, ut'. Mersenne also prepared a 'grand et immense' manuscript [45] that listed all 40 320 permutations of *eight* notes on 672 folio pages, followed by ranking and unranking algorithms. The important algorithm known as 'plain changes' (see [36, pp. 321–4]) was invented in England a few years later.

Methods for listing all permutations of a multiset with *repeated* elements were often misunderstood by early authors. For example, when Bhāskara listed the permutations of {4, 5, 5, 5, 8} in Section 271 of his *Līlāvatī* (*c.*1150), he gave them in the following order:

48555	84555	54855	58455	55485
55845	55548	55584	45855	45585
45558	85455	85545	85554	54585
58545	55458	55854	54558	58554 .

Mersenne used a slightly more sensible, but not completely systematic, order when he listed sixty anagrams of the Latin name IESVS [44, p. 131]. When Athanasius Kircher wanted to illustrate the thirty permutations of a five-note melody in [30, pp. 10–11], this lack of a system got him into trouble, as one is omitted and another appears twice.



But John Wallis knew better. In his *Discourse of Combinations* (1685) [69, pp. 117, 126], published as a supplement to his *Treatise of Algebra*, he correctly listed the sixty anagrams of 'messes' in lexicographic order (if we let $m < e < s$), and he recommended respecting alphabetical order 'that we may be the more sure, not to miss any'.

We will see later that the Indian mathematician Nārāyaṇa Paṇḍita had already developed a theory of permutation generation in the 14th century, although his work remained almost totally unknown.

Seki's list

Takakazu Seki (1642–1708) was a charismatic teacher and researcher who revolutionized the study of mathematics in 17th-century Japan. While he was studying the elimination of variables from simultaneous homogeneous equations, he was led to expressions such as

$$a_1b_2 - a_2b_1 \quad \text{and} \quad a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1,$$

which we now recognize as *determinants.*

In 1683 he published a booklet about this discovery, introducing an ingenious scheme for listing all permutations in such a way that half of them were 'alive' (even) and the other half were 'dead' (odd). Starting with the case $n = 2$, when '12' was alive and '21' was dead, he formulated the following rules for $n > 2$:

1. *Take every live permutation for* $n - 1$, *increase all of its elements by* 1, *and insert* 1 *in front. This produces* $(n-1)!/2$ *'basic permutations' of* $\{1, 2, \ldots, n\}$.
2. *From each basic permutation, form* $2n$ *others by rotation and reflection:*

$$a_1a_2 \ldots a_{n-1}a_n,\ a_2 \ldots a_{n-1}a_na_1,\ \ldots,\ a_na_1a_2 \ldots a_{n-1};$$

$$a_na_{n-1} \ldots a_2a_1,\ a_1a_na_{n-1} \ldots a_2,\ \ldots,\ a_{n-1} \ldots a_2a_1a_n.$$

If n *is odd, those in the first row are alive and those in the second are dead; if* n *is even, those in each row are alternatively alive, dead, … , alive, dead.*

For example, when $n = 3$, the only basic permutation is 1 2 3 . Thus 1 2 3, 2 3 1, 3 1 2 are alive while 3 2 1, 1 3 2, 2 1 3 are dead, and we have successfully generated the six terms of a 3×3 determinant. The basic permutations for

$n = 4$ are 1 2 3 4, 1 3 4 2, and 1 4 2 3, and from (say) 1 3 4 2 we get a set of eight – namely,

$$+1\,3\,4\,2 - 3\,4\,2\,1 + 4\,2\,1\,3 - 2\,1\,3\,4 + 2\,4\,3\,1 - 1\,2\,4\,3 + 3\,1\,2\,4 - 4\,3\,1\,2,$$

which are alternately alive (+) and dead (−). A 4×4 determinant therefore includes the terms $a_1b_3c_4d_2 - a_3b_4c_2d_1 + \cdots - a_4b_3c_1d_2$.

Seki's rule for permutation generation is quite pretty, but unfortunately it has a serious problem: it doesn't work when $n > 4$. His error seems to have gone unrecognized for hundreds of years (see [46]).

Lists of combinations

The earliest exhaustive list of *combinations* known to have survived the ravages of time appears in the last book of Suśruta's well-known Sanskrit treatise on medicine, Chapter 63, written before AD 600 and perhaps much earlier. Noting that medicine can be sweet, sour, salty, pungent, bitter, or astringent, Suśruta's book diligently listed the (15, 20, 15, 6, 1, 6) cases that arise when those qualities occur two, three, four, five, six, and one at a time.

Bhāskara repeated this example in Sections 110–14 of *Līlāvatī*, and observed that the same reasoning applies to six-syllable poetic metres with a given number of long syllables, but he simply mentioned the totals (6, 15, 20, 15, 6, 1) without listing the combinations themselves. In Sections 274 and 275, he observed that the numbers

$$\frac{n \times (n-1) \times \cdots \times (n-k+1)}{k \times (k-1) \times \cdots \times 1}$$

enumerate *compositions* (that is, ordered partitions) as well as combinations; again he gave no list.

> To avoid prolixity this is treated in a brief manner;
> for the science of calculation is an ocean without bounds.
>
> Bhāskara (c.1150)

As we see in Chapter 3, an isolated but interesting list of combinations appeared in 1144 in the remarkable algebra text *Kitāb al-bāhir* (Flamboyant Book), written by as-Samaw'al of Baghdad when he was only 19 years old. In the closing part of that work he presented a list of $C(10, 6) = 210$ simultaneous linear equations in ten unknowns, given here in modern notation:

$$(1)\ x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 65,$$

$$(2)\ x_1 + x_2 + x_3 + x_4 + x_5 + x_7 = 70,$$

$$(3)\ x_1 + x_2 + x_3 + x_4 + x_5 + x_8 = 75,$$

$$\vdots$$

$$(209)\ x_4 + x_6 + x_7 + x_8 + x_9 + x_{10} = 91,$$

$$(210)\ x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} = 100.$$

Each combination of ten things taken six at a time yielded one of his equations. His purpose was evidently to demonstrate that overdetermined equations can still have a unique solution – which in this case is

$$(x_1, x_2, \ldots, x_{10}) = (1, 4, 9, 16, 25, 10, 15, 20, 25, 5)$$

(see [3]).

Rolling dice

Some glimmerings of elementary combinatorics arose also in medieval Europe, especially in connection with the question of listing all possible outcomes when three dice are thrown. There are $C(8, 3) = 56$ ways to choose three objects from six when repetitions are allowed. Gambling was officially prohibited, yet these fifty-six ways became rather well known.

In about AD 965 Bishop Wibold of Cambrai in northern France devised a game called *Ludus Clericalis* (clerical game) so that members of the clergy could enjoy rolling dice while remaining pious. His idea was to associate each possible roll of three dice with one of fifty-six virtues; for example a roll of three 1s corresponded to *love* (*caritas*), the best virtue of all. Players took turns, and the first to roll each virtue acquired it. After all possibilities had arisen, the most virtuous player won. Wibold gave a complicated scoring system by which two virtues could be combined if the sum of the pips on all six of their dice was 21; for example, *love* + *humility* (1 1 1 + 6 6 6) or *chastity* + *intelligence* (1 2 4 + 4 5 5) could be paired in this way, and such combinations ranked above any individual virtue. He also considered more complex variants of the game in which vowels appeared on the dice instead of spots, so that virtues could be claimed if their vowels were thrown.

Wibold's list of fifty-six virtues was presented in lexicographic order, from 1 1 1 to 1 1 2 to $\cdots$ to 6 6 6, when it was first described by Baldéric in his *Chronicon Cameracense* (Chronicle of Cambrai), about 150 years later [5]. But another medieval manuscript presented the possible dice rolls in quite a different order:

666	555	444	333	222	111	665
664	663	662	661	556	554	553
552	551	446	445	443	442	441
336	335	334	332	331	226	225
224	223	221	116	115	114	113
112	654	543	432	321	643	641
631	531	653	652	651	621	521
421	542	541	643	642	532	431.

In this case the author knew how to deal systematically with repeated values, but had a complicated ad hoc way to handle the twenty cases in which all three dice were different, so he listed 643 twice and missed 632 (see [7]).

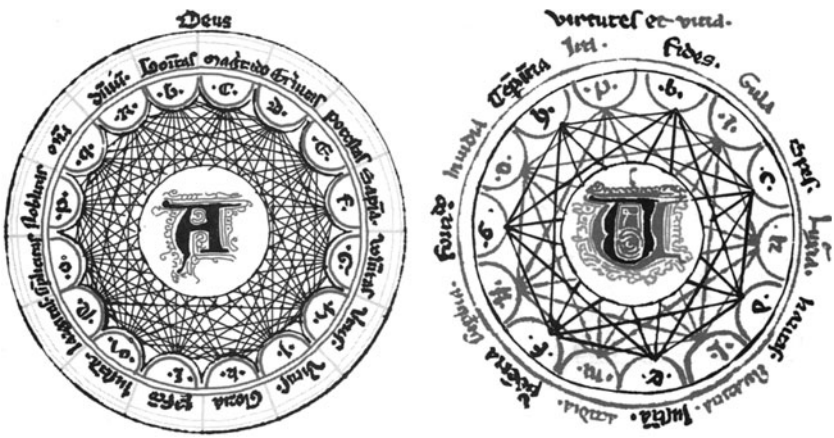
An amusing poem entitled *The Chaunce of the Dyse*, attributed to John Lydgate, was written in the early 1400s for use at parties. Its opening verses invite each person to throw three dice; then the remaining verses, which are indexed in decreasing lexicographic order from 666 to 665 to ... to 111, give fifty-six character sketches that light-heartedly describe the thrower (see [24]).

> I pray to god that euery wight may caste
> Vpon three dyse ryght as is in hys herte
> Whether he be rechelesse or stedfaste
> So moote he lawghen outher elles smerte
> He that is gilty his lyfe to converte
> They that in trouthe haue suffred many a throwe
> Moote ther chaunce fal as they moote be knowe.
>
> *The Chaunce of the Dyse* (c.1410)

Ramon Llull

Significant ripples of combinatorial concepts also emanated from an energetic and quixotic Catalan poet, novelist, encyclopedist, educator, mystic, and missionary, named Ramon Llull (*c*.1232–1316). As we see in Chapter 5, Llull's approach to knowledge was essentially to identify basic principles and then to contemplate combining them in all possible ways.

For example, one chapter in his *Ars Compendiosa Inveniendi Veritatem* (The Concise Art of Finding the Truth) (*c*.1274) began by enumerating sixteen attributes of God: *goodness, greatness, eternity, power, wisdom, love, virtue, truth, glory, perfection, justice, generosity, mercy, humility, sovereignty*, and *patience*. Then Llull wrote $C(16,2) = 120$ short essays of about 80 words each, considering God's *goodness* as related to *greatness*, God's *goodness* as related to *eternity*, and so on, ending with God's *sovereignty* as related to *patience*. In another chapter he considered seven virtues (*faith, hope, charity, justice, prudence, fortitude, temperance*) and seven vices (the 'seven deadly sins': *gluttony, lust, greed, sloth, pride, envy, anger*), with $C(14,2) = 91$ subchapters to deal with each pair in turn. Other chapters were systematically divided in a similar way, into $C(8,2) = 28$, $C(15,2) = 105$, $C(4,2) = 6$, and $C(16,2) = 120$ subsections. (One wonders what might have happened if he had been familiar with Wibold's list of 56 virtues: would he have produced commentaries on all $C(56,2) = 1540$ of their pairs?)



Illustrations in a manuscript presented by Ramon Llull to the doge of Venice in 1280.

Llull illustrated his methodology by drawing circular diagrams like those above (see [40]). The left-hand circle in this illustration, *Deus*, names sixteen divine *attributes* – essentially the same sixteen listed earlier, except that *love* (*amor*) was now called *will* (*voluntas*), and the final four were now *simplicity, rank, mercy*, and *sovereignty*, in that order. Each attribute was assigned a code letter, and the illustration depicts their interrelations as the complete graph K_{16} on vertices $B, C, D, E, F, G, H, I, K, L, M, N, O, P, Q, R$. The right-hand figure, *virtutes et vitia*, shows the seven virtues (b, c, d, e, f, g, h) interleaved with the seven vices (i, k, l, m, n, o, p); in the original manuscript virtues appeared in blue ink, while vices appeared in red. Notice that in this case his illustration depicted two independent complete graphs K_7, one of each colour. (He no longer bothered to compare each individual virtue with each individual vice, since every virtue was clearly better than every vice.)

Llull used the same approach to write about medicine. Instead of juxtaposing theological concepts, his *Liber Principiorum Medicinæ* (Book of the Principles of Medicine) (*c*.1275) considered combinations of symptoms and treatments. He also wrote books on philosophy, logic, jurisprudence, astrology, zoology, geometry, rhetoric, and chivalry – more than 200 works in all. It must be admitted, however, that much of this material was highly repetitive; modern data compression techniques would probably reduce Llull's output to a size much less than that of (say) Aristotle.

A Llullian illustration from a manuscript presented to the queen of France around 1325.

He eventually decided to simplify his system by working primarily with groups of nine things. For instance, he first listed only the first nine of God's attributes ($B, C, D, E, F, G, H, I, K$) in a circle ($BC, BD, \ldots, IK$) (see [39]). The $C(9, 2) = 36$ associated pairs then form a staircase diagram that he placed to the right of that circle. By adding two more virtues (*patience* and *compassion*) as well as two more vices (*lying* and *inconsistency*) he could treat virtues vis-à-vis virtues and vices vis-à-vis vices with the same chart. He also proposed using the same chart to carry out an interesting scheme for voting, in an election with nine candidates (see [43]).

Three encircled triangles in Llull's chart illustrate another key aspect of his approach. A triangle (B, C, D) stands for (*difference, concordance, contrariness*); a triangle (E, F, G) stands for (*beginning, middle, ending*); and a triangle (H, I, K) stands for (*greater, equal, less*). These three interleaved appearances of K_3 represent three kinds of three-valued logic. Llull had experimented earlier with other such triplets – notably (*true, unknown, false*).

We can get an idea of how he used the triangles by considering how he dealt with combinations of the four basic elements (*earth, air, fire, water*). All four elements are different; *earth* is concordant with *fire*, which concords with *air*, which concords with *water*, which concords with *earth*; *earth* is contrary to *air*, and *fire* is contrary to *water*; these considerations complete an analysis with respect to triangle (B, C, D). Turning to triangle (E, F, G), he noted that various processes in nature begin with one element dominating another; then a transition or middle state occurs, until a goal is reached, like air becoming warm. For triangle (H, I, K) he said that in general we have *fire* > *air* > *water* > *earth* with respect to their 'spheres', their 'velocities', and their 'nobilities'; nevertheless we also have, for example, *air* > *fire* with respect to supporting *life*, while *air* and *fire* have equal value when they are working together.

Llull provided a vertical table to the right of his chart as a further aid. He also introduced movable concentric wheels, labelled with the letters (B,C,D,E,F,G,H,I,K) and with other names, so that many things could be contemplated simultaneously. In this way a faithful practitioner of the Llullian art could be sure to have all the bases covered.

Several centuries later, in 1669, Kircher published an extension of Llull's system as part of a large tome entitled *Ars Magna Sciendi Sive Combinatoria* (The Great Art of Knowledge, or the Combinatorial Art) [31] (see frontispiece), with five movable wheels accompanying page 173 of that book. Kircher also extended Llull's repertoire of complete graphs by providing illustrations of complete bipartite graphs.

It is an investigative and inventive art. When ideas are combined in all possible ways, the new combinations start the mind thinking along novel channels and one is led to discover fresh truths and arguments.

Martin Gardner, *Logic Machines and Diagrams* (1958)

The most extensive modern development of Llull-like methods is perhaps *The Schillinger System of Musical Composition* (1946) by Joseph Schillinger [57], a remarkable two-volume work that presents theories of rhythm, melody, harmony, counterpoint, composition, orchestration, etc., from a combinatorial perspective. On page 56, for example, Schillinger lists the twenty-four permutations of $\{a, b, c, d\}$ in the Gray-code order of plain changes; then on page 57 he applies them not to pitches but rather to rhythms, to the durations of notes. On page 364 he exhibits the symmetrical cycle

$$(2, 0, 3, 4, 2, 5, 6, 4, 0, 1, 6, 2, 3, 1, 4, 5, 3, 6, 0, 5, 1),$$

a universal cycle of 2-combinations for the seven objects $\{0, 1, 2, 3, 4, 5, 6\}$. This is an Eulerian trail in K_7 : all $C(7, 2) = 21$ pairs of digits occur exactly once. Such patterns are grist to a composer's mill. But we can be grateful that Schillinger's better students (such as George Gershwin) did not commit themselves entirely to a strictly mathematical sense of aesthetics.

Tacquet, van Schooten, and Izquierdo

Three additional books related to our story were published during the 1650s. André Tacquet wrote a popular text, *Arithmeticæ Theoria et Praxis* [66], that was reprinted and revised often during the next fifty years. Near the end, on pages 376 and 377, he gave a procedure for listing combinations two at a time, then three at a time, etc.

Frans van Schooten's *Exercitationes Mathematicæ* [58] was more advanced. On page 373 he listed all combinations in an appealing layout,

$$\begin{array}{c} a \\ \hline b.\ ab \\ \hline c.\ ac.\ bc.\ abc \\ \hline d.\ ad.\ bd.\ abd.\ cd.\ acd.\ bcd.\ abcd \end{array},$$

and he proceeded on the next few pages to extend this pattern to the letters e, f, g, h, i, k, 'et sic in infinitum'. On page 376 he observed that one can replace (a, b, c, d) by $(2, 3, 5, 7)$ to get the divisors of 210 that exceed unity:

$$\begin{array}{c} 2 \\ \hline 3\ \ 6 \\ \hline 5\ \ 10\ \ 15\ \ 30 \\ \hline 7\ \ 14\ \ 21\ \ 42\ \ 35\ \ 70\ \ 105\ \ 210 \end{array}.$$

On the following page he extended the idea to

$$\begin{array}{c} \underline{\;a\;} \\ \underline{a.\ aa} \\ \underline{b.\ ab.\ aab} \\ c.\ ac.\ aac.\ bc.\ abc.\ aabc \end{array},$$

thereby allowing two *a*s. He did not really understand this extension, though; his next example

$$\begin{array}{c} \underline{\;a\;} \\ \underline{a.\ aa} \\ \underline{a.\ aaa} \\ \underline{b.\ ab.\ aab.\ aaab} \\ b.\ bb.\ abb.\ aabb.\ aaabb \end{array}$$

was botched, indicating the limits of his knowledge at the time. On page 411 van Schooten observed that the weights $(a, b, c, d) = (1, 2, 4, 8)$ could be assigned in his first layout above, leading to

$$\begin{array}{c} \underline{\;1\;} \\ \underline{2\ \ 3} \\ \underline{4\ \ 5\ \ 6\ \ 7} \\ 8\ \ 9\ \ 10\ \ 11\ \ 12\ \ 13\ \ 14\ \ 15 \end{array}$$

after addition, but he failed to see the connection with base-2 arithmetic.

Sebastián Izquierdo's two-volume work *Pharus Scientiarum* (Lighthouse of Sciences) [27] included a nicely organized discussion of combinatorics entitled Disputatio 29, *De Combinatione*. He gave a detailed discussion of four key parts of Stanley's Twelvefold Way [62, Sec. 1.4] – namely, the n-tuples, n-variations, n-multi-combinations, and n-combinations of m objects.

In Sections 81–84 of *De Combinatione* he listed all combinations of m letters taken n at a time, for $2 \leq n \leq 5$ and $n \leq m \leq 9$, always in lexicographic order; he also tabulated them for $m = 10$ and 20 in the cases $n = 2$ and 3. But when he listed the $m(m-1)\cdots(m-n+1)$ *variations* of m things taken n at a time, he chose a more complicated ordering [36, Ex. 7.2.1.7–14].

Izquierdo was first to discover the formula $C(m+n-1, n)$ for combinations of m things taken n at a time with unlimited repetition; this rule appeared in Sections 48–51 of his work. But in Section 105, when he attempted to list all such combinations in the case $n = 3$, he was unaware that there was a simple way to do it. In fact, his listing of the fifty-six cases for $m = 6$ was rather like the old awkward ordering of dice rolls that we saw above.

Combinations with repetition were not well understood until Jacob Bernoulli's *Ars Conjectandi* (The Art of Conjecturing) came out in 1713 (see Chapter 6). In Part 2, Chapter 5, Bernoulli simply listed the possibilities in lexicographic order, and showed that the formula $C(m+n-1, n)$ follows by induction as an easy consequence. Niccolò Tartaglia had, incidentally, come close to discovering this formula in 1556 in his *General Trattato di Numeri, et Misure* (General Treatise of Numbers and Measures) [67]; so had the Maghreb mathematician Ibn Mun^{c}im in his 13th-century *Fiqh al-Ḥisāb* (see Chapter 3).

The null case

Before we conclude our discussion of early work on combinations, we should not forget a small yet noble step taken by John Wallis on page 110 of his *Discourse of Combinations* [69], where he specifically considered the combination of m things taken 0 at a time:

> It is manifest, That, if we would *take None*, that is, if we would *leave All*; there can be but one case thereof, what ever be the Number of things exposed.

Furthermore, on page 113, he knew that $C(0, 0) = 1$:

> for, here, to take all, or to leave all, is but one and the same case.

However, when he gave a table of $n!$ for $n \leq 24$, he did not go so far as to point out that $0! = 1$, or that there is exactly one permutation of the empty set.

The work of Nārāyaṇa

A remarkable monograph entitled *Gaṇitakaumudī* (Lotus Delight of Calculation), written by Nārāyaṇa Paṇḍita in 1356, has recently become known in detail to scholars outside of India for the first time, thanks to an English translation by Parmanand Singh [61] (see Chapter 1). Chapter 13 of Nārāyaṇa's work, subtitled *Aṅka Pāśa* (Concatenation of Numbers), was devoted to combinatorial generation. Indeed, although the ninety-seven 'sutras' of this chapter were rather cryptic, they presented a comprehensive theory of the subject that anticipated developments in the rest of the world by several hundred years.

For example, Nārāyaṇa dealt with permutation generation in sutras 49–55a, where he gave algorithms to list all permutations of a set in decreasing colexicographic order, together with algorithms to rank a given permutation and to unrank a given serial number. These algorithms had appeared more than a century earlier in the well-known work *Saṅgītaratnākara* (Jewel-Mine of Music) by Śārṅgadeva, Sections 1.4.60–71, who thereby had essentially discovered the factorial representation of positive integers. Nārāyaṇa went on in sutras 57–60 to extend Śārṅgadeva's algorithms so that general multisets could readily be permuted; for example, he listed the permutations of $\{1, 1, 2, 4\}$ as

$$1124, 1214, 2114, 1142, 1412, 4112, 1241, 2141, 1421, 4121, 2411, 4211,$$

again in decreasing colexicographic order.

Nārāyaṇa's sutras 88–92 dealt with systematic generation of combinations. Besides illustrating the combinations of $\{1, 2, \ldots, 8\}$ taken three at a time – namely,

$$(678,\ 578,\ 478,\ \ldots,\ 134,\ 124,\ 123)$$

– he also considered a bit-string representation of these combinations in the reverse order (*increasing* colexicographic order), extending a 10th-century method of Bhaṭṭotpala:

$$(11100000,\ 11010000,\ 10110000,\ \ldots, 00010011,\ 00001011,\ 00000111).$$

Thus we can legitimately regard Nārāyaṇa Paṇḍita as the founder of the science of combinatorial generation – even though, like many other pioneers who were significantly 'ahead of their time', his work on the subject never became well known, even in his own country.

Permutable poetry

Let us turn now to a curious question that attracted the attention of several prominent mathematicians in the 17th century, because it sheds considerable light on the state of combinatorial knowledge in Europe at that time. A Jesuit priest named Bernard Bauhuis [6] had composed a famous one-line tribute to the Virgin Mary, in Latin hexameter:

Tot tibi sunt dotes, Virgo, quot sidera cælo.
(Thou hast as many virtues, O Virgin, as there are stars in heaven.)

His verse inspired Erycius Puteanus, a professor at the University of Louvain, to write a book entitled *Pietatis Thaumata* (Miracles of Piety) [55], presenting 1022 permutations of Bauhuis's words. For example, Puteanus wrote:

107 Tot dotes tibi, quot cælo sunt sidera, Virgo.
270 Dotes tot, cælo sunt sidera quot, tibi Virgo.
329 Dotes, cælo sunt quot sidera, Virgo tibi tot.
384 Sidera quot cælo, tot sunt Virgo tibi dotes.
725 Quot cælo sunt sidera, tot Virgo tibi dotes.
949 Sunt dotes Virgo, quot sidera, tot tibi cælo.
1022 Sunt cælo tot Virgo tibi, quot sidera, dotes.

He stopped at 1022, because 1022 was the number of visible stars in Ptolemy's well-known catalogue of the heavens.

The idea of permuting words in this way was well known at the time; such word play was what Julius Scaliger had called 'Proteus verses' in his *Poetices Libri Septem* (Seven Books of Poetry) [56]. The Latin language lends itself to permutations, because Latin word endings tend to define the function of each noun, making the relative word order much less important to the meaning of a sentence than it is in English. Puteanus did state, however, that he had specifically avoided unsuitable permutations such as

Sidera tot cælo, Virgo, quot sunt tibi dotes,

because they would place an *upper* bound on the Virgin's virtues, rather than a lower bound (see pp. 12, 103 of his book).

Of course, there are $8! = 40\,320$ ways to permute the words of 'Tot tibi sunt dotes, Virgo, quot sidera cælo'. But that wasn't the point; most of those ways don't 'scan'. Each of Puteanus's 1022 verses obeyed the strict rules of classical *hexameter*, the rules that had been followed by Greek and Latin poets since the days of Homer and Vergil – namely:

(i) each word consists of syllables that are either long ($-$) or short ($\smile$);
(ii) the syllables of each line belong to one of 32 patterns,

$$\left\{\begin{matrix}-\smile\smile\\--\end{matrix}\right\}\left\{\begin{matrix}-\smile\smile\\--\end{matrix}\right\}\left\{\begin{matrix}-\smile\smile\\--\end{matrix}\right\}\left\{\begin{matrix}-\smile\smile\\--\end{matrix}\right\}-\smile\smile\left\{\begin{matrix}-\smile\\--\end{matrix}\right\};$$

in other words there are six metrical feet, where each of the first four is either a dactyl (—⌣⌣) or a spondee (——), the fifth foot should be a dactyl, and the last is either a trochee (—⌣) or a spondee.

The rules for long versus short syllables in Latin poetry are somewhat tricky in general, but the eight words of Bauhuis's verse can be characterized by the following patterns:

$$\text{tot} = -\,,\ \ \text{tibi} = \left\{\begin{matrix}\smile\smile\\ \smile\,-\end{matrix}\right\},\ \ \text{sunt} = -\,,\ \ \text{dotes} = --\,,$$

$$\text{Virgo} = \left\{\begin{matrix}-\,\smile\\ --\end{matrix}\right\},\ \ \text{quot} = -\,,\ \ \text{sidera} = -\smile\smile,\ \text{and cælo} = --\,.$$

Notice that poets had two choices when they used the word 'tibi' or 'Virgo'. Thus, for example, the line from Bauhuis's verse fits the hexameter pattern

— ⌣ ⌣	— —	— —	— —	— ⌣ ⌣	— —
Tot ti- bi	sunt do-	tes, Vir-	go, quot	si-de- ra	cæ-lo.

(Dactyl, spondee, spondee, spondee, dactyl, spondee: 'dum-diddy dum-dum dum-dum dum-dum dum-diddy dum-dum'. The commas represent slight pauses, called 'cæsuras', when the words are read; they do not concern us here, although Puteanus inserted them carefully into each of his 1022 permutations.)

A natural question now arises: if we permute Bauhuis's words at random, what are the odds that they scan? In other words, how many of the permutations obey rules (i) and (ii), given the syllable patterns listed above? Leibniz raised this question, among others, in his *Dissertatio de Arte Combinatoria* (1666), a work published when he was applying for a position at the University of Leipzig (see Chapter 6). At this time Leibniz was just 19 years old, largely self-taught, and his understanding of combinatorics was quite limited; for example, he believed that there are 600 permutations of {ut, ut, re, mi, fa, sol} and 480 of {ut, ut, re, re, mi, fa}, and he even stated that rule (ii) represents 76 possibilities instead of 32.

But Leibniz did realize that it would be worthwhile to develop general methods for counting all permutations that are 'useful', in situations when many permutations are 'useless'. He considered several examples of Proteus verses, enumerating some of the simpler ones correctly but making many errors when the words were complicated. Although he mentioned Puteanus's work, he didn't attempt to enumerate the scannable permutations of Bauhuis's famous line.

A much more successful approach was introduced a few years later by Jean Prestet in his 1675 *Elémens des Mathématiques* [53]. Prestet gave a clear exposition leading to the conclusion that exactly 2196 permutations of Bauhuis's verse would yield a proper hexameter. However, he soon realized that he had forgotten to count quite a few cases – including those numbered 270, 384, and 725 in the list given above. So he completely rewrote this material when he published his *Nouveaux Elémens des Mathématiques* in 1689. Pages 127–33 of Prestet's new book were devoted to showing that the true number of scannable permutations was 3276, almost 50% larger than his previous total.

Meanwhile John Wallis had treated the problem in his *Discourse of Combinations* [69, pp. 118–19]. After explaining why he believed the correct number to be 3096, Wallis admitted that he may have overlooked some possibilities or counted some cases more than once: 'but I do not, at present, discern either the one and other'.

An anonymous reviewer of Wallis's work remarked that the true number of metrically correct permutations is actually 2580 – but he gave no proof [1]. This reviewer was almost certainly Leibniz himself, although no clue to the reasoning behind the number 2580 has been found among Leibniz's voluminous unpublished notes.

Finally, Jacob Bernoulli entered the picture. In his inaugural lecture as Dean of Philosophy at the University of Basel in 1692, he mentioned the tot-tibi enumeration problem and stated that a careful analysis is necessary to obtain the correct answer – which, he said, was 3312(!). His proof appeared posthumously in the first edition of his *Ars Conjectandi* [8]. Bernoulli did not actually intend to publish those pages in this now famous book, but the proofreader who found them among his notes decided to include the full details, in order 'to gratify curiosity' (see [9]).

So who was right? Are there 2196 scannable permutations, or 3276, or 3096, or 2580, or 3312? W. A. Whitworth and W. E. Hartley considered the question anew in 1902 in *The Mathematical Gazette* [71], where they each presented elegant arguments and concluded that the true total is in fact none of the above. Their joint answer, 2880, represented the first time that any two mathematicians had independently come to the same conclusion about this problem.

But, in the end, Bernoulli was vindicated, and everybody else was wrong (see [36, Ex. 7.2.1.7–21]). Moreover, a study of Bernoulli's systematic and carefully indented three-page derivation indicates that he was successful chiefly because he adhered faithfully to a discipline that we now call the *backtrack method*.

Even the wisest and most prudent people often suffer from what Logicians call insufficient enumeration of cases.

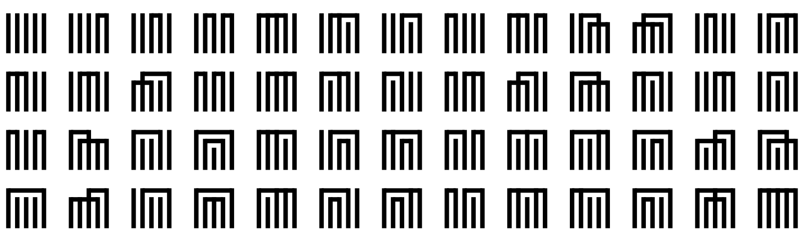
Jacob Bernoulli (1692)

Set partitions

The partitions of a set seem to have been studied first in Japan, where a parlour game called *genji-ko* (Genji Incense) became popular among upperclass people around AD 1500. The host of a gathering would secretly select five packets of incense, some of which might be identical, and he would burn them one at a time. The guests would try to discern which of the scents were the same and which were different; in other words, they would try to guess which of the $\varpi_5 = 52$ set partitions of $\{1, 2, 3, 4, 5\}$ had been chosen by their host.

Diagrams used to represent set partitions in 16th-century Japan.

Soon it became customary to represent the fifty-two possible outcomes by diagrams such as these. For example, the first diagram above, when read from right to left, would indicate that the first two scents are identical and so are the last three; thus the partition is 1 2 | 3 4 5; the other two diagrams, again read from right to left, are pictorial ways to represent the respective partitions 1 2 4 | 3 5 and 1 | 2 4 | 3 5. As an aid to memory, each of the fifty-two patterns was named after a chapter of Lady Murasaki's famous 11th-century *Tale of Genji*, according to the following sequence (see [19]):



(Once again, as we have seen in many other examples, the possibilities were not arranged in any particularly logical order.)

The appealing nature of these genji-ko patterns led many families to adopt them as heraldic crests. For example, stylized variants of the fifty-two patterns were found in standard catalogues of kimono patterns early in the 20th century (see [2]).

Early in the 1700s Takakazu Seki and his students began to investigate the number of set partitions ϖ_n for arbitrary n, inspired by the known result that $\varpi_5 = 52$. Yoshisuke Matsunaga found formulas for the number of set partitions when there are k_j subsets of size n_j, for $1 \le j \le t$, with $k_1 n_1 + k_2 n_2 + \cdots + k_t n_t = n$. He also discovered the basic recurrence relation

$$\varpi_{n+1} = C(n,0)\,\varpi_n + C(n,1)\,\varpi_{n-1} + C(n,2)\,\varpi_{n-2} + \cdots + C(n,n)\,\varpi_0,$$

from which the values of ϖ_n can readily be computed.

Matsunaga's discoveries remained unpublished until Arima Yoriyuki's book *Shūki Sanpō* (The Essences of Mathematics) came out in 1769. Problem 56 of that book asked the reader to solve the equation $\varpi_n = 678\,570$ for n; Yoriyuki's answer, worked out in detail (with credit duly given to Matsunaga), was $n = 11$.

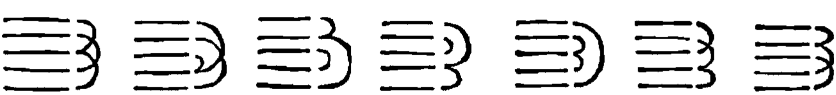
Shortly afterwards, Masanobu Saka studied the number $\left\{ {n \atop k} \right\}$ of ways that an n-set can be partitioned into k subsets. In his work *Sanpō-Gakkai* (The Sea of Learning on Mathematics) (1782), he discovered the recurrence formula

$$\left\{ {n+1 \atop k} \right\} = k\left\{ {n \atop k} \right\} + \left\{ {n \atop k-1} \right\},$$

and tabulated the results for $n \le 11$. James Stirling, in his *Methodus Differentialis* (Method of Differentials) (1730), had discovered the numbers $\left\{ {n \atop k} \right\}$ in a purely algebraic context; thus Saka was the first person to realize their combinatorial significance. These numbers are now known as *Stirling numbers of the second kind*, often denoted by $S(n,k)$ (see Chapter 6).

An interesting algorithm for listing set partitions was subsequently devised by Toshiaki Honda (see [36, Ex. 7.2.1.7–24]). Further details about genji-ko and its relation to the history of mathematics can be found in two Japanese articles by Tamaki Yano [73].

Set partitions remained virtually unknown in Europe until much later, except for three isolated incidents. First, George Puttenham published *The Arte of English Poesie* in 1589, and pages 70–2 of that book contain diagrams similar to those of genji-ko. For example, the following seven diagrams were used to illustrate possible rhyme schemes for five-line poems, 'whereof some of them be harsher and unpleasaunter to the eare then other some be'. But his visually appealing list was incomplete.



Second, an unpublished manuscript of Leibniz from the late 1600s shows that he had tried to count the number of ways to partition $\{1, 2, \ldots, n\}$ into three or four subsets, but with almost no success. He enumerated ${n \brace 2}$ by a very cumbersome method, which would not have led him to see readily that ${n \brace 2} = 2^{n-1} - 1$. He attempted to compute ${n \brace 3}$ and ${n \brace 4}$ only for $n \leq 5$, and made several numerical slips leading to incorrect answers (see [32, pp. 229–33] and [34, pp. 316–21]).

The third European appearance of set partitions had a completely different character. John Wallis devoted the third chapter of his *Discourse of Combinations* (1685) to questions about 'aliquot parts', the proper divisors of numbers, and in particular he studied the set of all ways to factorize a given integer. This question is equivalent to the study of *multiset* partitions; for example, the factorizations of p^3q^2r are essentially the same as the partitions of $\{p, p, p, q, q, r\}$, when p, q, and r are prime numbers. Wallis devised an excellent algorithm for listing all factorizations of a given integer n, but he did not investigate the important special cases that arise when n is the power of a prime (equivalent to integer partitions) or when n is squarefree (equivalent to set partitions). Thus, although Wallis was able to solve the more general problem, its complexities paradoxically deflected him from discovering partition numbers, Bell numbers, or Stirling subset numbers, or from devising simple algorithms that would generate integer partitions or set partitions.

Integer partitions

Partitions of integers arrived on the scene even more slowly (see Chapter 9). We saw above that Bishop Wibold (*c.*965) knew the partitions of n into exactly three parts not exceeding 6. So did Galileo [22], who wrote a memo about them (*c.*1627), listing partitions in decreasing lexicographic order, and also studied their frequency of occurrence as rolls of three dice. Thomas Strode extended this to four dice [64], and Thomas Harriot, in unpublished work a few years earlier, had considered up to six dice (see [63]).

Mersenne listed the partitions of 9 into any number of parts on page 130 of his 1636 *Traitez de la Voix et des Chants* (Treatise on the Voice and Singing). For each partition $9 = a_1 + a_2 + \cdots + a_k$ he also computed the multinomial coefficient $9!/(a_1!a_2! \cdots a_k!)$; as we have seen earlier, he was interested in counting various melodies, and he knew (for example) that there are $9!/3!\,3!\,3! = 1680$

melodies on the nine notes $\{a, a, a, b, b, b, c, c, c\}$. But he failed to mention the cases $8 + 1$ and $3 + 2 + 1 + 1 + 1 + 1$, probably because he had not listed the possibilities in any systematic way.

Leibniz considered two-part partitions in Problem 3 of his 1666 *Dissertatio de Arte Combinatoria* (see Chapter 6), and his unpublished notes show that he subsequently spent considerable time trying to enumerate the partitions with three or more summands. He called them 'discerptions' or (less frequently) 'divulsions' – in Latin, of course – or sometimes 'sections' or 'dispersions' or even 'partitions'. He was interested in them primarily because of their connection with the monomial symmetric functions $\sum x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots$. But his many attempts led to almost total failure, except in the case of three summands, where he almost (but not quite) discovered a general formula. For example, he carelessly counted only twenty-one partitions of 8, forgetting the case $2 + 2 + 2 + 1 + 1$; and he got only 26 for $p(9)$, after missing $3 + 2 + 2 + 2$, $3 + 2 + 2 + 1 + 1$, $2 + 2 + 2 + 1 + 1 + 1$, and $2 + 2 + 1 + 1 + 1 + 1 + 1$ – in spite of the fact that he was trying to list partitions systematically in decreasing lexicographic order (see [32, pp. 91–258], [33, pp. 409–30], and [34, pp. 255–337]).

Abraham de Moivre had the first real success with partitions in 1697, in his paper 'A method of raising an infinite multinomial to any given power, or extracting any given root of the same' [47]. He proved that the coefficient of z^{m+n} in $(az + bz^2 + cz^3 + \cdots)^m$ has one term for each partition of n; for example, the coefficient of z^{m+6} is

$$\begin{aligned} &C(m,6)a^{m-6}b^6 + 5C(m,5)a^{m-5}b^4c + 4C(m,4)a^{m-4}b^3d \\ &\quad + 6C(m,4)a^{m-4}b^2c^2 + 3C(m,3)a^{m-3}b^2e + 6C(m,3)a^{m-3}bcd \\ &\quad + 2C(m,2)a^{m-2}bf + C(m,3)a^{m-3}c^3 + 2C(m,2)a^{m-2}ce \\ &\quad + C(m,2)a^{m-2}d^2 + C(m,1)a^{m-1}g. \end{aligned}$$

If we set $a = 1$, the term with exponents $b^i c^j d^k e^l \ldots$ corresponds to the partition with i 1s, j 2s, k 3s, l 4s, etc. Thus, for example, when $n = 6$ he essentially presented the partitions in the order

111111, 11112, 1113, 1122, 114, 123, 15, 222, 24, 33, 6.

He explained how to list the partitions recursively, as follows (but in different language related to his own notation): for $k = 1, 2, \ldots, n$, start with k and append the (previously listed) partitions of $n - k$ whose smallest part is at least k.

[My solution] was ordered to be published in the Transactions, not so much as a matter relating to Play, but as containing some general Speculations not unworthy to be considered by the Lovers of Truth.

Abraham de Moivre (1717)

Pierre Rémond de Montmort tabulated all partitions of numbers not exceeding 9 into 6 or fewer parts in his *Essay d'Analyse sur les Jeux de Hazard* (1708), in connection with dice problems (see Chapters 6 and 13). His partitions were listed in a different order from de Moivre; for example,

$$111111,\ 21111,\ 2211,\ 222,\ 3111,\ 321,\ 33,\ 411,\ 42,\ 51,\ 6.$$

He was probably unaware of de Moivre's prior work.

So far almost none of the authors we have been discussing actually bothered to describe the procedures by which they generated combinatorial patterns. We can only infer their methods, or lack thereof, by studying the lists that they actually published. Furthermore, in rare cases such as de Moivre's paper where a tabulation method *was* explicitly described, the author assumed that all patterns for the first cases $1, 2, \dots, n-1$ had been listed before it was time to tackle the case of order n. No method for generating patterns 'on the fly', moving directly from one pattern to its successor without looking at auxiliary tables, was actually explained by any of the authors we have encountered, except for Kedāra and Nārāyaṇa. Today's computer programmers naturally prefer methods that are more direct and need little memory.

In 1747 Roger Joseph Boscovich [10] published the first direct algorithm for partition generation. His method produces for $n = 6$ the respective outputs

$$111111,\ 11112,\ 1122,\ 222,\ 1113,\ 123,\ 33,\ 114,\ 24,\ 15,\ 6.$$

As it happens, precisely the reverse order turns out to be slightly easier and faster than the order that he had chosen (see [36, pp. 391–2]).

Boscovich published sequels in 1748 [11], extending his algorithm in two ways. First, he considered generating only partitions whose parts belong to a given set S, so that symbolic multinomials with sparse coefficients could be raised to the mth power. (He said that the greatest common divisor of all elements of S should be 1; in fact, however, his method could fail if $1 \notin S$.) Second, he introduced an algorithm for generating partitions of n into m parts, given m and n. Again he was unlucky: a slightly better way to do that task was found subsequently, diminishing his chances for fame (see [36, pp. 392–3]).

Hindenburg's hype

The inventor of that better way for generating partitions of n into m parts was Carl Friedrich Hindenburg (see Chapter 12), who also rediscovered Nārāyaṇa's algorithm for generating multiset permutations. Unfortunately, these small successes led him to believe that he had made revolutionary advances in mathematics – although he did condescend to remark that other people such as de Moivre, Euler, and Lambert had come close to making similar discoveries.

Hindenburg was a prototypical over-achiever, extremely energetic if not inspired. He founded or co-founded Germany's first professional journals of mathematics (published in 1786–9 and 1794–1800), and contributed long articles to each. He served several times as academic dean at the University of Leipzig, where he was also the Rector in 1792. Had he been a better mathematician, German mathematics might well have flourished more in Leipzig than in Berlin or Göttingen.

But his first mathematical work, *Beschreibung einer ganz neuen Art, nach einem bekannten Gesetze fortgehende Zahlen durch Abzählen oder Abmessen bequem und sicher zu finden* [26], amply foreshadowed what was to come: his 'ganz neuen Art' (completely new art) idea in that booklet was simply to give combinatorial significance to the digits of numbers written in decimal notation. Incredibly, he concluded his monograph with large foldout sheets that contained a table of the numbers from 0000 through 9999 – followed by two other tables that listed the even numbers and odd numbers separately!

Hindenburg published letters from people who praised his work, and he invited them to contribute to his journals. In 1796 he edited *Sammlung combinatorisch–analytischer Abhandlungen*, whose subtitle stated (in German) that de Moivre's multinomial theorem was 'the most important proposition in all of mathematical analysis'. About a dozen people joined forces to form what became known as *Hindenburg's Combinatorial School*, and they published thousands of pages filled with esoteric symbolism that must have impressed many non-mathematicians.

The work of this School was not completely trivial from the standpoint of computer science. For example, H. A. Rothe, who was Hindenburg's best student, noticed that there is a simple way to go from a Morse code sequence to its lexicographic successor or predecessor. Another student, J. C. Burkhardt, observed that Morse code sequences of length n could also be generated easily, by first considering those with no dashes, then one dash, then two, etc. Their motivation was not to tabulate poetic metres of n beats, as it had been in India, but rather to list the terms of the continuant polynomials $K(x_1, x_2, \ldots, x_n)$ (see [4]). Furthermore, on page 53 of Hindenburg's 1796 *Sammlung* (Collection) cited above, G. S. Klügel introduced a way to list all permutations that has subsequently become known as Ord-Smith's algorithm (see [36, pp. 330–1]).

Hindenburg believed that his methods deserved equal time with algebra, geometry, and calculus in the standard curriculum. But he and his disciples were combinatorialists who only made combinatorial lists. Burying themselves in formulas and formalisms, they rarely discovered any new mathematics of real interest. Eugen Netto [51] has admirably summarized their work:

> For a while they controlled the German market; however, most of what they dug up soon sank into a not-entirely-deserved oblivion.

The sad outcome was that combinatorial studies in general got a bad name. Gösta Mittag-Leffler, who assembled a magnificent library of mathematical literature about one hundred years after Hindenburg's death, decided to place all such work on a special shelf marked 'Dekadenter' (decadent). And this category still persists in the library of Sweden's Institut Mittag-Leffler today, even as that institute attracts world-class combinatorial mathematicians whose research is anything but decadent.

Looking on the bright side, we may note that at least one good book did emerge from all of this activity. Andreas von Ettingshausen's 1826 *Die combinatorische Analysis* [21] is noteworthy as the first text to discuss combinatorial generation methods in a perspicuous way. He discussed the general principles of lexicographic generation in Section 8, and applied them to construct good ways to list all permutations (Section 11), combinations (Section 30), and partitions (Sections 41–4).

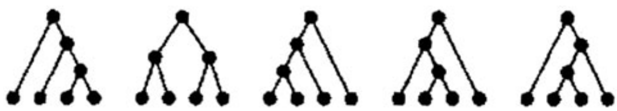
Where were the trees?

We have now seen that lists of n-tuples, permutations, combinations, and partitions were compiled rather early in recorded history, by interested and interesting researchers. That covers almost all of the important combinatorial objects that frequently need to be listed, except for the various kinds of tree structures. Thus, our story will be complete if we can trace the origins of tree generation (see also Chapters 8 and 12).

But the historical record of that topic before the advent of computers is virtually a blank page, with the exception of a few 19th-century papers by Arthur Cayley. Cayley's major work on trees [14], originally published in 1875, was climaxed by a large foldout illustration that exhibited all of the free (unrooted) trees with nine or fewer unlabelled vertices. Earlier in that paper he had also illustrated the nine oriented (rooted) trees with five vertices. The methods he used to produce those lists were quite complicated. All free trees with up to ten vertices were listed many years later by F. Harary and G. Prins [25], who also went up to $n = 12$ in the cases of free trees with no vertices of degree 2 or with no symmetries.

The trees most dearly beloved by computer scientists – binary trees, or the equivalent ordered forests or nested parentheses – are strangely absent from the literature. Many mathematicians of the 1700s and 1800s had learned how to count binary trees, and we also know that the Catalan numbers C_n enumerate dozens of different kinds of combinatorial objects. Yet before 1950 nobody seems to have published an actual *list* of the $C_4 = 14$ objects of order 4 in *any* of these guises, much less the $C_5 = 42$ objects of order 5, except indirectly: out of a total of fifty-two genji-ko diagrams (see above), the forty-two that have no intersecting lines turn out to be equivalent to the five-vertex binary trees and forests, but this fact was not realized until the 20th century.

There are a few isolated instances where authors of yore did prepare lists of $C_3 = 5$ Catalan-related objects. Cayley, again, was first; in [13] he illustrated the binary trees with three internal vertices and four leaves as follows:



(That same paper also illustrated another species of tree, equivalent to so-called weak orderings.) Then, in 1901, Eugen Netto [50] listed the five ways to insert parentheses into the expression $a + b + c + d$:

$$(a+b)+(c+d),\quad \bigl((a+b)+c\bigr)+d,\quad \bigl(a+(b+c)\bigr)+d,$$
$$a+\bigl((b+c)+d\bigr),\quad a+\bigl(b+(c+d)\bigr).$$

The five permutations of $\{+1,+1,+1,-1,-1,-1\}$ whose partial sums are nonnegative were listed in the following way by Paul Erdős and Irving Kaplansky [20]:

$$1+1+1-1-1-1,\quad 1+1-1+1-1-1,\quad 1+1-1-1+1-1,$$
$$1-1+1+1-1-1,\quad 1-1+1-1+1-1.$$

Even though only five objects are involved in each of these three examples, we can see that the orderings in the first two examples were basically catch-as-catch-can; only the last ordering was systematic and lexicographic.

We should also note briefly the work of Walther von Dyck, since many recent papers use the term 'Dyck words' to refer to strings of nested parentheses. Dyck was an educator known for co-founding the Deutsches Museum in Munich, among other things. He wrote two pioneering papers about the theory of free groups [18]. Yet the so-called Dyck words have at best a tenuous connection to his actual research: he studied the words on $\{x_1, x_1^{-1}, x_2, x_2^{-1}, \ldots, x_k, x_k^{-1}\}$ that reduce to the empty string after repeatedly erasing adjacent letter-pairs of the forms $x_i x_i^{-1}$ or $x_i^{-1} x_i$; the connection with parentheses and trees arises only when we limit erasures to the first case, $x_i x_i^{-1}$, and he never considered such a limitation.

Thus we may conclude that, although an explosion of interest in binary trees and their cousins occurred after 1950, such trees represent the only aspect of our story whose historical roots are rather shallow.

After 1950

The arrival of electronic computers changed everything. The first computer-oriented publication about combinatorial generation methods was a note in 1956 by C. B. Tompkins, 'Machine attacks on problems whose variables are permutations' [68]. Thousands more were destined to follow.

Several articles by D. H. Lehmer, especially his 'Teaching combinatorial tricks to a computer' [38], proved to be extremely influential in the early days. Lehmer represented an important link to previous generations. For example, Stanford University's library records show that he had checked out Netto's *Lehrbuch der Combinatorik* in January 1932.

The main publications relevant to particular algorithms of modern importance are cited in [36], so there is no need to repeat them here. But textbooks and monographs that first put pieces of the subject together in a coherent framework were also of great importance. Three books, in particular, were especially noteworthy with respect to establishing general principles:

- *Elements of Combinatorial Computing* by Mark B. Wells (Pergamon Press, 1971), especially Chapter 5.
- *Combinatorial Algorithms* by Albert Nijenhuis and Herbert S. Wilf (Academic Press, 1975); a second edition was published in 1978, containing additional material, and Wilf subsequently wrote *Combinatorial Algorithms: An Update* (SIAM, 1989).
- *Combinatorial Algorithms: Theory and Practice* by Edward M. Reingold, Jurg Nievergelt, and Narsingh Deo (Prentice-Hall, 1977), especially Chapter 5.

Robert Sedgewick compiled the first extensive survey of permutation generation methods in *Computing Surveys* 9 (1977), 137–64, 314. Carla Savage's survey article about Gray codes in *SIAM Review* 39 (1997), 605–29, was another milestone.

We noted above that algorithms to generate Catalan-counted objects were not invented until computer programmers developed an appetite for them. The first such algorithms to be published have been superseded by better techniques, but it is appropriate to list them here. First, H. I. Scoins [59] gave two recursive algorithms for ordered tree generation. His algorithms dealt with binary trees represented as bit strings that were essentially equivalent to Polish prefix notation or to nested parentheses. Then Mark Wells, in Section 5.5.4 of his book cited above, generated binary trees by representing them as non-crossing set partitions, and Gary Knott [35] gave recursive ranking and unranking algorithms for binary trees.

Algorithms to generate all spanning trees of a given graph have been published by numerous authors ever since the 1950s, motivated originally by the study of electrical networks. Among the earliest such papers were works of N. Nakagawa [48], W. Mayeda [42], H. Watanabe [70], and S. Hakimi [23].

A recent introduction to the entire subject can be found in Chapters 2 and 3 of *Combinatorial Algorithms: Generation, Enumeration, and Search* by Donald L. Kreher and Douglas R. Stinson [37].

Frank Ruskey has been preparing a book entitled *Combinatorial Generation* that contains a thorough treatment and a comprehensive bibliography.

References

1. *Acta Eruditorum* 5 (1686), 289.
2. F. Adachi, *Japanese Design Motifs*, Dover (1972), 150–3.
3. S. Ahmad and R. Rashed, *Al-Bāhir en Algèbre d'As-Samaw'al*, Damascus (1972), 77–82.

4. *Archiv für reine und angewandte Mathematik* 1 (1794), 154–94.
5. Baldéric le Rouge, *Chronicon Cameracense*, *Patrologia Latina 134*, Paris (1884), 1007–16.
6. B. Bauhuis, *Epigrammatum Libri V*, Cologne (1615), 49 (also p. 47 of his *Epigrammatum Selectorum*, Antwerp, 1618, 1620, …) .
7. D. R. Bellhouse, *De Vetula*: a medieval manuscript containing probability calculations, *Internat. Stat. Rev.* 68 (2000), 123–36.
8. J. Bernoulli, *Ars Conjectandi* (1713), 79–81.
9. *Die Werke von Jakob Bernoulli*, Vol. 3, Birkhäuser, Basel (1975), 78, 98–106, 108, 154–5.
10. R. J. Boscovich, *Giornale de' Letterati*, Rome (1747), 393–404 (together with two foldout tables facing p. 404).
11. R. J. Boscovich, *Giornale de' Letterati*, Rome (1748), 12–27, 84–99.
12. C. P. Brown, *Sanskrit Prosody* (1869), 28.
13. A. Cayley, On the theory of the analytical forms called trees. Second part, *Philos. Mag.* 18 (1859), 374–8 = *Collected Mathematical Papers* 4, 112–15.
14. A. Cayley, On the analytical forms called trees, with application to the theory of chemical combinations, *Report of the British Association for the Advancement of Science* (1875), 257–305 = *Collected Mathematical Papers* 4, 427–60.
15. W. von Christ, *Metrik der Griechen und Römer* (1879), 78–9.
16. J. Drexel, *Orbis Phaëthon*, Munich (1629), 668–71.
17. J. Drexel, *Orbis Phaëthon*, Cologne edition (1631), 526–31.
18. W. von Dyck, Gruppentheoretische Studien, *Math. Ann.* 20 (1882), 1–44; 22 (1883), 70–108.
19. *Encyclopedia Japonicæ*, Sanseido, Tokyo (1910), 1299.
20. P. Erdős and I. Kaplansky, Sequences of plus and minus, *Scripta Math.* 12 (1946), 73–5.
21. A. von Ettingshausen, *Die Combinatorische Analysis*, Vienna (1826).
22. G. Galilei, Sopra le scoperte dei dadi, *Opere* (1898) 8, 591–4.
23. S. Hakimi, On trees of a graph and their generation, *J. Franklin Inst.* 272 (1961), 347–59.
24. E. P. Hammond, The chance of the dice, *Englische Studien* 59 (1925), 1–16.
25. F. Harary and G. Prins, The number of homeomorphically irreducible trees, and other species, *Acta Math.* 101 (1959), 141–62.
26. C. F. Hindenburg, *Beschreibung einer ganz neuen Art, nach einem bekannten Gesetze fortgehende Zahlen durch Abzählen oder Abmessen bequem und sicher zu finden*, Leipzig (1776).
27. S. Izquierdo, *Pharus Scientiarum* 2, Lyon (1659), 319–58.
28. S. Kak, Yamātārājabhānasalagām: an interesting combinatoric sūtra, *Indian J. Hist. Sci.* 35 (2000), 123–7.
29. H. Keil, *Grammatici Latini* 1 (1857), 474–82.
30. A. Kircher, *Musurgia Universalis* 2, Haeredes Francisci Corbelletti, Rome (1650), 32.

31. A. Kircher, *Ars Magna Sciendi Sive Combinatoria*, Johannes Janssonius a Waesberge, Amsterdam (1669).
32. E. Knobloch, Die mathematischen Studien von G. W. Leibniz zur Kombinatorik, *Studia Leibnitiana Supplementa* 11 (1973).
33. E. Knobloch, The mathematical studies of G. W. Leibniz on combinatorics, *Historia Math.* 1 (1974), 409–30.
34. E. Knobloch, Die mathematischen Studien von G. W. Leibniz zur Kombinatorik: Textband, *Studia Leibnitiana Supplementa* 16 (1976).
35. G. Knott, A numbering system for binary trees, *Commun. ACM* 20 (1977), 113–15.
36. D. E. Knuth, *The Art of Computer Programming*, Vol. 4A, Addison-Wesley (2011).
37. D. L. Kreher and D. R. Stinson, *Combinatorial Algorithms: Generation, Enumeration, and Search*, CRC Press (1999).
38. D. H. Lehmer, Teaching combinatorial tricks to a computer, *Proc. Symp. Appl. Math.* 10 (1960), 179–93 (see also *Proc.1957 Canadian Math. Congress* (1959), 160–73; *Proc. IBM Scientific Computing Symposium on Combinatorial Problems* (1964), 23–30; and *Applied Combinatorial Mathematics* (ed. E. F. Beckenbach), Wiley (1964), Ch. 1, 5–31).
39. Llullian illustrations from a manuscript presented to the queen of France, *c.*1325, Badische Landesbibliothek Karlsruhe, Codex St. Peter perg. 92, folios 28^{v} and 39^{v}.
40. R. Llull, *Ars Demonstrativa*, Biblioteca Marciana, VI 200, folio 3^{v} (1280).
41. R. J. Lynn, *The Classic of Changes*, Columbia University Press (1994).
42. W. Mayeda, Reducing computation time in the analysis of networks by digital computer, *IRE Trans.* CT-6 (1959), 136–7, 394.
43. I. McLean and J. London, Ramon Lull and the theory of voting, *Studia Lulliana* 32 (1992), 21–37.
44. M. Mersenne, *Traitez de la Voix et des Chants* (Vol. 2 of *Harmonie Universelle*, 1636), 111–15, 117–28.
45. M. Mersenne, Bibliothèque Nationale de France, Fonds Français, no. 24256.
46. Y. Mikami, *The Development of Mathematics in China and Japan* (1913), 191–9; *Takakazu Seki's Collected Works*, Osaka (1974), 18–20.
47. A. de Moivre, A method of raising an infinite multinomial to any given power, or extracting any given root of the same, *Philos. Trans.* 19 (1697), 619–25 and Fig. 5.
48. N. Nakagawa, On evaluation of the graph trees and the driving point admittance, *IRE Trans.* CT-5 (1958), 122–7.
49. J. Needham, *Science and Civilisation in China*, Vol. 2, Cambridge University Press (1956), 304–45.
50. E. Netto, *Lehrbuch der Combinatorik* (1901), Sec. 122.52/53: *Elemens.*
51. E. Netto, in M. Cantor's *Geschichte der Mathematik* 4 (1908), 201–19.
52. B. van Nooten, Binary numbers in Indian antiquity, *J. Indian Philos.* 21 (1993), 31–50.
53. J. Prestet, *Elémens des Mathématiques*, Paris (1675), 342–438.
54. J. Prestet, *Nouveaux Elémens des Mathématiques*, Paris (1689), 127–33.
55. E. Puteanus, *Pietatis Thaumata*, Antwerp (1617).
56. J. Scaliger, *Poetices Libri Septem*, Book 2, Lyon (1561), Ch. 30, p. 73.

57. J. Schillinger, *The Schillinger System of Musical Composition*, Carl Fischer (1946).
58. F. van Schooten, *Exercitationes Mathematicæ*, Johannes Elzevier, Leiden (1657).
59. H. I. Scoins, Placing trees in lexicographic order, *Mach. Intell.* 3 (1968), 43–60.
60. P. Singh, The so-called Fibonacci numbers in ancient and medieval India, *Historia Math.* 12 (1985), 229–44.
61. P. Singh, The Gaṇita Kaumudī of Nārāyaṇa Paṇḍita, *Gaṇita Bhāratī* 20 (1998), 25–82; 21 (1999), 10–73; 22 (2000), 19–85; 23 (2001), 18–82; 24 (2002), 35–98. (See also the PhD thesis of T. Kusuba, Brown University, 1993.)
62. R. P. Stanley, *Enumerative Combinatorics*, Vol. 1, Wadsworth (1986).
63. J. Stedall, Symbolism, combinations, and visual imagery in the mathematics of Thomas Harriot, *Historia Math.* 34 (2007), 380–401, on p. 398.
64. T. Strode, *Short Treatise of the Combinations, Elections, Permutations and Composition of Quantities*, London (1678), 19–20.
65. F. Swetz, Leibniz, the Yijing, and the religious conversion of the Chinese, *Math. Mag.* 76 (2003), 276–91.
66. A. Tacquet, *Arithmeticæ Theoria et Praxis*, Louvain (1656).
67. N. Tartaglia, *General Trattato di Numeri, et Misure*, Vol. 2, Venice (1556), 17^{r} and 69^{v}.
68. C. B. Tompkins, Machine attacks on problems whose variables are permutations, *Proc. Symp. Appl. Math.* 6 (1956), 202–5.
69. J. Wallis, *Discourse of Combinations*, London (1685).
70. H. Watanabe, A computational model for network topology, *IRE Trans.* CT-7 (1960), 296–302.
71. W. A. Whitworth and W. E. Hartley, The number of ways of writing "Tot tibi sunt dotes, Virgo, quot sidera coelo," without disobeying the laws of metre, caesura excepted, *Math. Gazette* 2 (1902), 227–8.
72. Yang Hsiung, *Canon of Supreme Mystery* (transl. M. Nylan), Albany (1993).
73. T. Yano, Genjikoh and partitioning finite sets (in Japanese), *Sugaku Seminar* 34 (11) (Nov. 1995), 58–61; 34 (12) (Dec. 1995), 56–60.

This page intentionally left blank

PART II

Ancient Combinatorics

The arrangement of iconographic attributes in the multiple hands of deities, as in this depiction of a statue of Viṣṇu (Vishnu) holding his discus, conch, lotus, and mace, is a frequent theme of sample problems concerning permutations in Sanskrit texts.

CHAPTER 1

Indian combinatorics

TAKANORI KUSUBA AND KIM PLOFKER

From ancient times Indian scholars have shown interest in arranging things in regular order, applying ordering techniques to various types of items and concepts and theorizing about such arrangements mathematically. This interest was manifested in rules for permutations, combinations, and enumeration, as well as in the study of series. In some areas, such as metrics and music, this knowledge may go back to before the beginning of the present era.

Combinations and permutations in ancient sources

Many early Sanskrit texts consider various possibilities for selecting and ordering elements in a given set of items. As early as the late Vedic period in the first millennium BC, texts known as *prātisākhyas* prescribed systematic ways of rearranging the syllables of the Vedic invocations. Since reciting the words of the sacred hymns flawlessly was considered crucial to the success of the accompanying sacrifices, Vedic priests memorized them – not only in their proper 'form', but also with their syllables reversed or otherwise re-ordered – to serve as a check on possible corruption of the oral tradition.

Lists of the possible results of selecting and ordering different elements in specified ways appear in a wide variety of canonical texts from the early Classical Sanskrit period; most cannot be precisely 'dated', but all are generally ascribed to the time span between the last few centuries BC and the first few centuries AD. For instance, the *Mānava-dharmaśāstra*, a seminal work on *dharma* or religious laws and practices, enumerates and critiques the various possible forms of 'mixed caste', arising from marriages between men and women belonging to each of the four standard *varṇas* or strata of society (Brāhmaṇa, Kṣatriya, Vaiśya, and Śūdra). Similarly, the well-known sexology treatise *Kāmasūtra* divides men and women into three categories, each based on their physical characteristics, and classifies the resulting nine possible heterosexual pairings.

But the *śāstra*, or science, in which permutations and combinations seem to have made their first appearance as the subject of generalized computational techniques, possibly also in the early Classical period, is *chandas* or Sanskrit prosody (metrics). In succeeding centuries the relevant applications of combinatorial rules were also mentioned in other *śāstras*, such as music theory, medicine, and architecture. Not until after the middle of the first millennium AD or thereabouts does the subject seem to have penetrated to treatises on mathematics or computation per se (Sanskrit *gaṇita*), eventually acquiring the specialized label *aṅkapāśa*, or 'net of digits'.

Metrics

Sanskrit literary composition is in the form of prose or verse, the latter genre going back at least as far as the earliest known Vedic hymns. Verse formats are regulated either by the number of syllables (*akṣara*) or by the number of syllabic instants (*mātrā* or mora) that make up the canonical 'foot' or *pāda* in a particular metre. (The word *pāda* means literally 'quarter', as one verse usually consists of four identical *pādas*.) The basic units in Sanskrit prosody are syllables with one *mātrā*/mora, called *laghu* 'light', and syllables with two moras, called *guru* 'heavy'. A light syllable is indicated in prosody texts by a stroke (here denoted by I) and a heavy syllable is indicated by a little curve (denoted by S). The following list shows the possible metric variations for a set of four syllables, followed by those for a set of six moras.

1.	SSSS	5.	SSIS	9.	SSSI	13.	SSII
2.	ISSS	6.	ISIS	10.	ISSI	14.	ISII
3.	SISS	7.	SIIS	11.	SISI	15.	SIII
4.	IISS	8.	IIIS	12.	IISI	16.	IIII

1.	SSS	5.	IIIIS	9.	SSII	13.	IIIIII
2.	IISS	6.	ISSI	10.	IISII		
3.	ISIS	7.	SISI	11.	ISIII		
4.	SIIS	8.	IIISI	12.	SIIII		

A standard book, regarded as authoritative in Sanskrit metrics, is the *Chandaḥsūtra* (Prosody Rules) of Piṅgala, perhaps composed about 200 BC. Beginning before the 10th century AD, other texts or commentaries on metrics treated the subject more or less in accordance with the *Chandaḥsūtra*'s presentation. Since Piṅgala's work is very brief and difficult to understand on its own, these commentaries have been used to support the following exposition of it. (See the detailed discussions in [1], [14], [10, pp. 61–74], and [16] for the original *sūtras* and various commentaries on them.)

Concerning the identification and arrangement of such variations, there are six fundamental notions (*pratyaya*) in the *Chandaḥsūtra*; the standard order in which later commentators treat them is slightly different:

prastāra (extension): a unique standard sequence of all possible variations derived from a given set of elements, as well as the method of producing the sequence;

naṣṭa (lost): a technique for determining an unknown or 'lost' variation corresponding to a given serial number in the *prastāra* sequence;

uddiṣṭa (indicated): the converse of the preceding – that is, a technique for determining the serial number of a given variation within its *prastāra*;

saṃkhyā (enumeration): determining the number of variations;

adhvan (way): the process of calculating the amount of space on a writing surface required for writing out a given *prastāra* (which we do not discuss);

lagukriyā (light and heavy [syllable] calculation): the procedure for finding the number of possible variations of a verse metre containing a given number of *laghu* or *guru* syllables.

Piṅgala expounded the fundamental procedures for the *pratyayas*, as applied to syllabic (non-moric) metres, in sixteen concise *sūtras* or aphorisms, most of which are translated and explained in the following section.

The basic *pratyayas*

Piṅgala's rule for *prastāra* or extension, given in *Chandaḥsūtra* 8.20–23, takes the form of a very brief example for the case of finding variations of three syllables:

20. heavy and light are [placed] in two ways;
21. there are two adjacent [syllables, heavy and light];
22. [in two ways] the adjacent heavy and light are [placed] separately;
23. there are eight [variations] with three [syllables].

The rule is explained more generally by a commentator as follows, and is illustrated by the *prastāra* for four syllables (shown in the table above):

- Begin with the variation where all syllables are heavy (SSSS).
- Then change the leftmost syllable to a light one to form the second variation. The sequence of possible choices for this first syllable is thus {S, I}.
- For the third variation, change only the second syllable in the original variation to light, and for the fourth, make the first syllable light as well.
  The sequence of possibilities for the first two syllables is now {SS, IS, SI, II}.
- Continue this process recursively, so that each variation with the *k*th syllable light, but the following syllables heavy, conforms to the standard sequence for the previous $k - 1$ syllables. The final variation contains only light syllables.

As the table above indicates, this system of extension for the possible metres with n syllables corresponds to the binary representation of the non-negative integers up to $2^n - 1$; if a heavy syllable represents 0 and a light syllable 1, and if the least significant digit is considered to occupy the leftmost place, then the *prastāra* shown is exactly equivalent to writing down the numbers from 0 to 15 in binary notation.

The algorithm for determining *naṣṭa*, or a 'lost' variation, is stated by Piṅgala as follows:

24. when [an even number] is halved, [write] light;
25. when [an odd number] is increased by 1, [divide by 2 and write] heavy.

We can interpret the rules as follows. If the serial number of the lost variation is even, it is halved and a light-syllable symbol is written down. If it is odd, it is increased by 1 and then halved, and a symbol for a heavy syllable is written. The process is continued until the number of heavy and light symbols shown is equal to the number of syllables in the variation.

For example, if we wish to find the eleventh variation in the *prastāra* for four-syllable metres, we would produce the symbol sequence shown below as follows. Add 1 to 11 because 11 is odd, and halve the sum to get 6. A symbol for a heavy syllable is recorded. Since 6 is even, it is halved to produce 3, denoted by a light symbol. 3 is odd, so the sum of 3 and 1 is halved, yielding 2 with a heavy symbol; 2 is even, so it is halved and a light symbol is recorded. The four symbols now transcribed correspond to the four syllables of the metre, so the procedure is complete and the desired eleventh variation is heavy–light–heavy–light, or SISI, as expected.

11	6	3	2	1
	S	I	S	I

For finding the serial number of an *uddiṣṭa*, or 'indicated' variation, the *Chandaḥsūtra* states:

26. [Below the last letter of the indicated variation one should put the number 1.] A first light is multiplied by 2 in reverse order.
27. One should subtract 1 from that [if it is heavy].

Using the same example as above, but now treating the variation itself as given and the corresponding serial number as sought, we proceed as follows (and as shown below). Write down the given variation's sequence SISI, and place 1 below its rightmost symbol. Stepping through the symbols 'in reverse order', we multiply that 1 by 2 and record the product (2) beneath the heavy symbol immediately to the left of it. Doubling the 2 in turn yields 4, but since the current place's symbol is heavy, the product 4 is decreased by 1 to give 3, which corresponds to the light symbol in the next leftward place. Doubling 3 yields 6 for the next and final place to the left, and doubling 6 yields 12, but since the current place is again marked with a heavy symbol, the product is diminished by 1, giving $12 - 1 = 11$, the desired serial number of the given variation.

	S	I	S	I
11	6	3	2	1

Piṅgala's algorithm for finding the total number of possible n-syllable variations (*saṃkhyā*) also makes use of halving and doubling to compute the number 2^n:

28. When [an even number is] halved, two [is written];
29. when [an odd number is diminished by] 1, [write] 0;

30. where 0 [is written], doubled;

31. where halved, multiplied by itself.

The case with $n = 4$ produces the computation shown below. Working from left to right, the number of syllables 4, being even, is halved and 2 is recorded; then the even result $4/2 = 2$ is also halved and this step is again marked with 2. The new result $2/2 = 1$ is odd, so it is diminished by 1, which is indicated by writing down 0. The remainder from the subtraction step is 0, meaning that the reduction of the syllable number is completed. Now we start with the marker 1 on the right, and proceed from right to left, doubling the current number every time we encounter a 0 and squaring it at every 2, so the final total is $((1 \times 2)^2)^2 = 16$.

$\Rightarrow$	$4/2 = 2$	$2/2 = 1$	$1 - 1 = 0$	
	2	2	0	
16	4	2	1	$\Leftarrow$

This procedure reduces the number of operations required to compute 2^n by replacing doubling with squaring whenever possible. Its efficiency is evident upon noting that it involves essentially the same task as the problem of constructing any integer n from a starting value of 0 by either adding 1 to the current value or doubling it, with each operation repeated as often as necessary – that is, adding 1 to any integer exponent m corresponds to doubling 2^m, while doubling m is equivalent to squaring 2^m. While it would take n operations to compute 2^n if only doubling were used, it takes fewer than $2 \log_2 n$ operations if both doubling and squaring are permitted (a fact that is exploited in modern computation algorithms under the name 'left-to-right binary exponentiation').

The *lagukriyā*, or 'light and heavy calculation', for determining how many n-syllable metres contain a specified number of light or heavy syllables in a *pāda*, is described very concisely:

34. The full in front.

1	1	1	1	1
1	2	3	4	
1	3	6		
1	4			
1				

The *meru* for carrying out the *lagukriyā* for four-syllable metres.

This procedure requires the construction of a so-called *meru* figure, named after the fabulous mountain in Indian cosmology that stands at the centre of the earth's concentric rings of continents and seas. As illustrated above, where $n = 4$, the *meru* is framed in a half-square formed by a sequence of $n + 1$ repetitions of 1, written horizontally and vertically. Then the number in each successive cell, working outward from the corner to the 'front', is the sum of the number to the left of it and the one above it. This produces a figure identical to what we now call *Pascal's triangle*, with the same combinatorial implications. For example, the final diagonal sequence 1 4 6 4 1 informs us that there is one variation of a four-syllable *pāda* containing only heavy syllables, four variations with three heavy syllables and one light one, six with two heavy syllables and two light ones, and so forth.

A fountain at the Chennai Mathematical Institute commemorates the combinatorial algorithms of the ancient Indian scholar Piṅgala. It features numbers in what is now called the Fibonacci sequence.

Some other combinatorial elements of prosody

The *Chandaḥsūtra* also briefly addresses some applications of *pratyayas* to moric metres, particularly their *prastāra*, or list of possible variations. Again, the key concept is a recursive pattern of syllable change, specified by the following mnemonic (see [14, p. 233]):

> Two heavy [syllables, SS], a heavy at the end [IIS], [then] the middle [ISI], [then] the beginning [SII], and [all] light [syllables, IIII].

Additional syllables required by the given number of moras – as illustrated by the earlier example of the thirteen variations for a set of six moras – are initially set to heavy and then sequentially changed to light, working from right to left.

This systematic rule for the *prastāra* indicates that early prosodists also understood how to enumerate the metres with a given number of moras – another combinatorially interesting problem. For, the set of variations for a metre with n moras consists of all the variations for a metre with $n - 2$ moras with an additional heavy syllable (equal to two moras) appended to each, plus all the variations for a metre with $n - 1$ moras with an additional light syllable (that is, a single mora) appended to each. Since there is one possible variation (I) for a one-mora metre, and there are two variations (S, II) for a two-mora metre, the number of possible variations for three moras is $1 + 2 = 3$ (IS, SI, III), and the number of variations for four moras is $2 + 3 = 5$ (SS, IIS, ISI, SII, IIII), and so on. In general, the number of possible variations for a metre with n moras is the Fibonacci number F_{n+1}.

The earliest surviving detailed expositions of these combinatorial concepts in prosody were composed by authors of (and commentators on) *chandas* texts, beginning around the end of the 1st millennium AD ([1, transl., p. 20], [14, pp. 232–5], and [10, p. 60–1]). They are relevant to the development of Sanskrit prosody and also to its counterpart in vernacular or Prakrit texts, including the language of the sacred Jaina scriptures. In fact, some of the notable authors in this period were Jainas, including Hemacandra in the 12th century and the author of the *Prākṛtapaiṅgala* at the start of the 14th century. Since most of their combinatorial techniques were treated more fully and systematically in the later work of Nārāyaṇa Paṇḍita, as discussed below, we do not elaborate on them here.

المجتمع واحدا ثمّ ضرب الثلاثة فى الاثنين الباقيين [١] فاجتمع ستّة ،
و لكنّ ذلك لا يصحّ فى أكثر الصنوف وكأنّه وقع فى النسخة فساد
فأمّا الوضع فإنّه إذًا كان هكذا : < < < ا
و هو أن يكون مزاج السطر الأيمن | < < ب
بالإغباب واحدا من آخر و مزاج < | < ج
السطر الأوسط اثنين من نوع و اثنين | | < د
من آخر و مزاج الأيسر أربعة من ذا < < | ه
و أربعة من ذاك بحسب أزواج الزوج | < | و
فى مزاجات الأسطر ثمّ زيد فى الحساب < | | ز
المذكور أنّ ابتداء الصفّ إن كان بحصّة | | | ح
ثقيل نُقص منها قبلَ الضرب واحدٌ و إن كان الضرب فى حصّة ثقيل
نُقص من المبلغ واحدٌ حَصَلَ المطلوب من عدد رتبة الصفّ ؛ وكما أنّ
أبيات العربيّة تنقسم لنصفين بعروض و ضرب فإنّ أبيات أولئك تنقسم
لقسمين يسمّى كل واحد منها رِجلا [٢] و هكذا يسمّيها اليونانيّون ارجلا [٢] [٣]
ما يتركّب منه من الكلمات سلابى و الحروف بالصوت و عدمه و الطول
و القصر و التوسّط ؛ و ينقسم البيت لثلاث أرجل و لأربع و هو الأكثر
و ربّما زيد فى الوسط رجل خامسة و لا تكون مقفّاة و لكن إن كان
آخر الرجل الأولى و الثانية حرفا واحدا كالقافية و كذلك آخر الثالثة
و الرابعة أيضا حرفا واحدا سمّى هذا النوع " آرَلْ " و يجوز فى آخر.

(١) فى ز، وش : الباقية (٢) من ز ، و فى ش : رجل (٣ – ٣) بياض فى ش الرجل

In his work *India* the 11th-century Muslim mathematician al-Bīrūnī described the combinatorics of Sanskrit metrics for Arabic readers. Here he explains the arrangement of the eight possible three-syllable Sanskrit metres, where each syllable may be either heavy (<) or light (|).

Combinatorial rules in other *śāstras*

A number of texts in other Sanskrit technical disciplines echoed some of the prosodists' rules about determining the various arrangements of choices from given sets of items. Bharata's renowned *Nāṭyaśāstra* on dance, music, and drama, which like the *Chandaḥsūtra* probably dates from the late 1st millennium BC or within a few centuries afterwards, contains a chapter on *chandas* in which Bharata explicitly quotes Piṅgala's formula for the *prastāra* or extension. One of his chapters on music also addresses combinatorial problems in laying out and naming the variations of melodic sequences formed with the seven notes (*tāna*) of the Indian scale. A later music treatise, the 13th-century *Saṅgītaratnākara* of Śārṅgadeva, extends the methods of combination and permutation to classify sequences of specified numbers of rhythmic beats (*tāla*), as well as those of *tāna* or musical notes (see [8] and [10, pp. 98–137]).

Combinations are also relevant in the field of *āyurveda*, or medicine, particularly in the area of pharmacology. The *Carakasaṃhitā* and *Suśrutasaṃhitā*, both from approximately the early 1st millennium AD, treat in a rather elementary way the combinations of the six basic tastes (sweet, pungent, astringent, sour, saline, bitter) and of the three humours in Indian medical theory. For example, *Carakasaṃhitā* 1.26.15–22 painstakingly lists all the numbers of possible combinations of tastes, selecting from one to six tastes at a time:

> Sweet is combined with sour and so on; sour and so on [are combined] with the remaining [tastes] separately. These substances consisting of two tastes amount to fifteen. The combination of sweet combined with sour and so on separately is [combined] with the remaining [tastes] separately: thus [is combined the combination] of sour, of saline, and of pungent. Twenty substances consisting of three tastes are mentioned according to counting . . . But there is one [substance] having six tastes. Thus sixty-three substances are indicated by enumerating the tastes.

A later section similarly seeks to list the varieties of diseases caused by imbalances of two out of the three bodily humours (see [10, pp. 139–44]).

Combinatorial themes also appear occasionally in a wide range of other texts, including a section on metrics in the sacred narrative compilation *Agnipurāṇa* [6, p. 229]. Sometimes a Sanskrit text on architecture refers to a related technical term, such as the above-mentioned *meru*, but it is not yet fully understood what these signify in this context (see [10, pp. 144–6]). A far more extensive application of combinatorial concepts is found in the chapter on perfumery in the 6th-century *Bṛhatsaṃhitā* of Varāhamihira, a voluminous work on divination and miscellaneous other subjects [10, pp. 149–66]. Varāhamihira and his 10th-century commentator Utpala discussed the various fragrances that can be compounded by mixing any four from a selection of sixteen standard ingredients in varying ratios, where each ingredient may constitute one, two, three, or four parts of the whole.

For instance, Varāhamihira noted that, for a given set of four ingredients, the number of ways to combine them in the proportion 4 : 3 : 2 : 1 is given by $4 \times 3 \times 2 \times 1 = 24$, and for four such sets comprising all the sixteen different ingredients, the total number of substances is $24 \times 4 = 96$. He then observed that there are 1820 ways to pick such a set of four from the sixteen possibilities, and declared that the resulting total number of possible substances is $1820 \times 96 = 174\,720$. It was pointed out by Utpala (and [4]) that this is inaccurate: the number of possible sets of four (1820) should be multiplied by the number of possible ways to combine *each* set of four (24) to give the total number of perfumes as $1820 \times 24 = 43\,680$.

The derivation of the combination number 1820 itself employs an interesting technique, cryptically enunciated by Varāhamihira as follows:

> [A number] is combined with each preceding [number] that has passed except for the last place; they call [this] enumeration (*saṃkhyā*) . . .

Utpala's commentary used the example of choosing four out of sixteen ingredients, as illustrated below (where, for convenience, the corresponding combination notation $C(n, k)$ is listed next to each number). Write the numbers 1 to 16 upwards in a vertical column; in a second column next to it write 1. Then add the number 1 at the bottom in the first column to the number 2 above it, and write the result in the second column above the preceding number. Add this sum to the third number 3 from the bottom in the first column; write the result above the sum in the second column. Continue this process till the top of the column is reached, and repeat it for a third and a fourth column (as shown below), neglecting the final number in each column. The last number in the fourth column is the number of ways to select four from sixteen ingredients, which is 1820.

16	$C(16,1)$							
15	$C(15,1)$	120	$C(16,2)$					
14	$C(14,1)$	105	$C(15,2)$	560	$C(16,3)$			
13	$C(13,1)$	91	$C(14,2)$	455	$C(15,3)$	1820	$C(16,4)$
12	$C(12,1)$	78	$C(13,2)$	364	$C(14,3)$	1365	$C(15,4)$
11	$C(11,1)$	66	$C(12,2)$	286	$C(13,3)$	1001	$C(14,4)$
10	$C(10,1)$	55	$C(11,2)$	220	$C(12,3)$	715	$C(13,4)$
9	$C(9,1)$	45	$C(10,2)$	165	$C(11,3)$	495	$C(12,4)$
8	$C(8,1)$	36	$C(9,2)$	120	$C(10,3)$	330	$C(11,4)$
7	$C(7,1)$	28	$C(8,2)$	84	$C(9,3)$	210	$C(10,4)$
6	$C(6,1)$	21	$C(7,2)$	56	$C(8,3)$	126	$C(9,4)$
5	$C(5,1)$	15	$C(6,2)$	35	$C(7,3)$	70	$C(8,4)$
4	$C(4,1)$	10	$C(5,2)$	20	$C(6,3)$	35	$C(7,4)$
3	$C(3,1)$	6	$C(4,2)$	10	$C(5,3)$	15	$C(6,4)$
2	$C(2,1)$	3	$C(3,2)$	4	$C(4,3)$	5	$C(5,4)$
1	$C(1,1)$	1	$C(2,2)$	1	$C(3,3)$	1	$C(4,4)$

Varāhamihira's method for $C(n,k)$, shown for $n = 16$ and $k = 4$.

Other rules in this chapter of the *Bṛhatsaṃhitā* explain how to identify and count the various combinations, and how to use a specified diagram (actually a modified magic square of order 4 – see [5]) to represent the different possible substances.

Assimilation of combinatorics in mathematics texts

It took significantly longer for Sanskrit authors on *gaṇita* (or mathematics proper) to incorporate such rules on combinations and permutations fully into their own subject. Initially, this appropriation appears to have taken the form of simply adapting combinatorial formulas and examples from other disciplines and listing them with other miscellaneous rules and topics. The first known instance occurs in the mathematical astronomy text *Brāhmasphuṭasiddhānta*, composed by Brahmagupta in 628, which combines a fairly systematic treatment of astronomical calculations with a more erratic selection of chapters on various related topics, including general arithmetic and algebra, astronomical instruments, and *chandas* (see [3, pp. 357–8] and [6, p. 230]). This latter section was mentioned by the 11th-century polymath al-Bīrūnī in his book *India* [13, pp. 147–50], in his attempt to explain Sanskrit metrics for Arabic-literate Muslims, but its twenty verses have yet to be fully studied and explained.

Al-Bīrūnī illustrates the possible combinations of syllables in the four *pādas* of the moric metre called *skandha*.

The arithmetic text *Pāṭīgaṇita* of Śrīdhara, dating to probably the 8th or 9th century, briefly states a rule for enumerating the combinations of any number of tastes, from two up to the canonical maximum of six (see [11, p. 325] and [10, pp. 86–7]). The 9th-century Jaina mathematician Mahāvīra went into considerably more detail in his arithmetic treatise *Gaṇitasārasaṅgraha* (see [10, pp. 87–9] and [12, pp. 93–4, 108–9]). His chapter on 'mixture problems' deals with calculations on investments and profits, indeterminate equations or the 'pulverizer', series, and some combinatorial rules – including one described as a type of 'pulverizer' for counting combinations of k out of a set of n choices, which may be expressed in modern notation as follows:

$$C(n,k) = \frac{(n-k+1) \times \cdots \times (n-2) \times (n-1) \times n}{k \times (k-1) \times (k-2) \times \cdots \times 1}.$$

In this chapter Mahāvīra prescribed techniques for the *pratyayas*, equivalent to those of the classical prosodists, and in an earlier chapter he gave a generalized version of Piṅgala's *saṃkhyā* rule for computing r^n for any positive integers r and n.

In the 12th-century *Līlāvatī* of Bhāskara II, which became more or less the standard second-millennium Sanskrit arithmetic textbook, prosody rules were again included in the sections on 'mixture problems' (including investments, interest, etc.) and on series. However, after stating his version of the above formula for $C(n,k)$, Bhāskara explicitly pointed out that it is equally applicable to problems in different fields (see [2, pp. 106–7], [10, p. 89], and [11, pp. 187–8]):

Numbers beginning with and increasing by one [are set down] in reverse order [and respectively] divided by [the same numbers] standing in order. [Each] next [fraction] is to be multiplied by the previous, and the one next to it by that [product], [and so on]; [the results] are the [numbers of] variations of one, two, three, etc. [objects]. This is considered to be general. Its application in metrics among [those] who know it is given in the chapter on metrics [in the *Brāhmasphuṭasiddhānta*]; in architecture in the variations of window frames and the partial-*meru* [diagram]; and in medicine in the variations of tastes. These are not described for fear of over-extension [of this book]. (*Līlāvatī*, 112–14)

That is, after writing down the n fractions

$$\frac{n}{1}, \frac{n-1}{2}, \frac{n-2}{3}, \ldots, \frac{1}{n},$$

we obtain the number of combinations of any k out of n objects from the product

$$C(n,k) = \frac{n}{1} \times \frac{n-1}{2} \times \frac{n-2}{3} \times \cdots \times \frac{n-k+1}{k}.$$

Besides identifying the previous incarnations of combinations and permutations as applications of the same general methods, Bhāskara devoted the final section of the *Līlāvatī* to a new application of them under the technical term *aṅkapāśa* or 'net of digits' (see [10, pp. 90–5], [2, pp. 274–85], and [11, p. 191]). As its name suggests, this subject focuses on how to determine the numbers that can be created from a given set of digits (excluding 0), although some of the examples illustrate non-numerical variations such as the $n!$ different statues of a specified deity that can be created by rearranging n iconographic attributes (trident, conch shell, lotus, etc.) on its n hands. Treating a permutation of digits as an integer also allows *aṅkapāśa* to explore new problems, such as finding the arithmetical sum of all the integers produced from a given set of digits. Bhāskara's *aṅkapāśa* rules compute the following quantities:

- the total number $n!$ of permutations of n different objects, and the sum of all the integers thus produced if the objects are digits;
- the total number $n!/(n_1!\,n_2!\cdots)$ of permutations of n non-distinct digits, if one digit is repeated n_1 times, another n_2 times, etc. – and how to find their sum in each case;
- the total number $m!/(m-n)!$ of permutations of n distinct digits chosen from m possibilities;

- the total number of permutations of n digits that add up to a specified sum s, provided that $s < n + 9$. Bhāskara's formula is

$$\frac{s-1}{1} \times \frac{s-2}{2} \times \frac{s-3}{3} \times \cdots \times \frac{s-n+1}{n-1}.$$

The combinatorial expositions of Nārāyaṇa Paṇḍita

The culmination of combinatorics as a mathematical topic in Sanskrit appears to have been reached in the *Gaṇitakaumudī* (Lotus Delight of Calculation), composed by Nārāyaṇa Paṇḍita in 1356, and in the accompanying commentary supplied either by Nārāyaṇa himself or by another author. The thirteenth chapter of the work is devoted to *aṅkapāśa*, and the fourteenth is devoted to the earliest currently known comprehensive Sanskrit exposition of constructing magic squares of any order and with any given sum (see [7]). Here we focus mostly on surveying the structure and innovations of Nārāyaṇa's Chapter 13 (see [10] and [15] for complete translations with commentary).

While evidently following the example of Bhāskara's *Līlāvatī* in much of his subject selection and organization, Nārāyaṇa also drew on sources in music, metrics, medicine, and miscellaneous topics such as the *Bṛhatsaṃhitā*. He merged his borrowings and new results into an unprecedentedly complete and systematic treatment of combinatorial rules which, as he noted, is generally applicable in many fields (see [10, p. 320]):

> It is applied to dance and music, metrics, medicine, garland-making, and mathematics as well as architecture. Knowledge of these is indeed [acquired] by means of the net of numbers.

Nārāyaṇa broke down his exposition of *aṅkapāśa* into sections on 'sequences' (*paṅkti*), 'operations' (*karaṇa*), and the *pratyayas*. The combinatorial variations to which his rules apply are considered to be integers composed of numerical digits between 1 and 9, although for some of the *pratyayas* the digits can stand as symbols for non-numerical objects, such as musical notes.

Among Nārāyaṇa's twelve types of sequence are the following:

- *an increasing sequence* from 1 up to a given n, or $\{1, 2, 3, \ldots, n\}$;
- *a sequence without intervals*, or the n-digit integer with each digit equal to 1;
- *a sequence with separations*, or the digit 1 repeated n times;
- *an additive sequence* for a given number of terms n and a given number of addends q, where the first two terms are 1, the third is their sum 2, and each

successive term up to the $(n+1)$st is the sum of the previous q terms – for example, when $n = 7$ and $q = 3$, the additive sequence is $\{1, 1, 2, 4, 7, 13, 24, 44\}$, and when $q = 2$, it is the sequence of Fibonacci numbers $\{1, 1, 2, 3, \ldots, F_{n+1}\}$ (see [14, pp. 237–8]);
- *a geometric progression sequence* $\{1, q, q^2, q^3, \ldots, q^n\}$, for given numbers n and q.

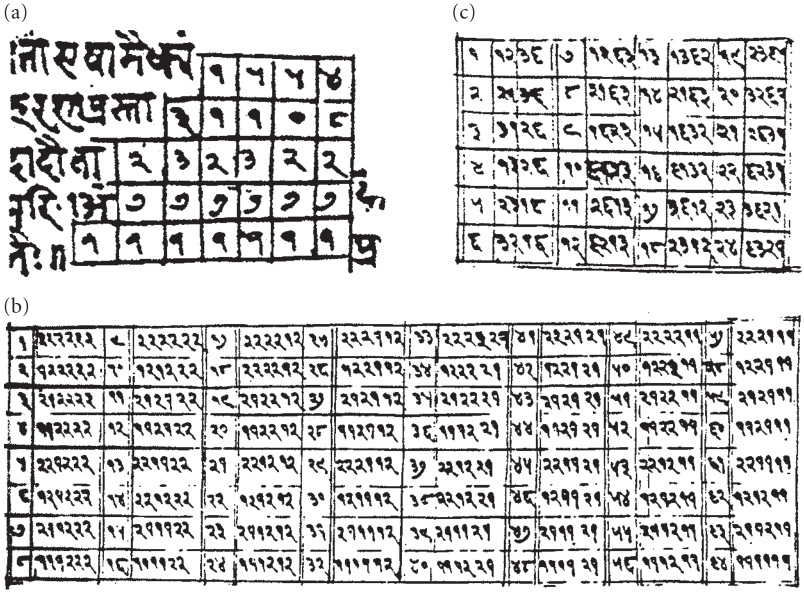
The 'operations' include tabular diagrams that are variations on the well-known *meru* figure illustrated above, and the use of a 'marker' to keep track of certain specified cells within a diagram. The *pratyayas* are fundamentally the same concepts enunciated in ancient prosody, but Nārāyaṇa supplemented them with many related problems and techniques to produce the following list of computable quantities:

- *saṃkhyā*, or the total number of variations;
- 'repetition', or how many times each of the given digits appears in a given place;
- the sum of the last digits in all the variations;
- the sum of all the variations;
- the total number of digits and total number of each distinct digit in all the variations;
- *prastāra*, or the standard list of all variations;
- the number of variations with a specified number of places;
- *naṣṭa*, or the particular variation corresponding to a given serial number within the *prastāra*;
- *uddiṣṭa*, or the serial number corresponding to a given variation;
- the number of variations whose digits add up to a given sum;
- the number of variations containing a specified number of repetitions of a given digit;
- the number of variations ending with a given digit.

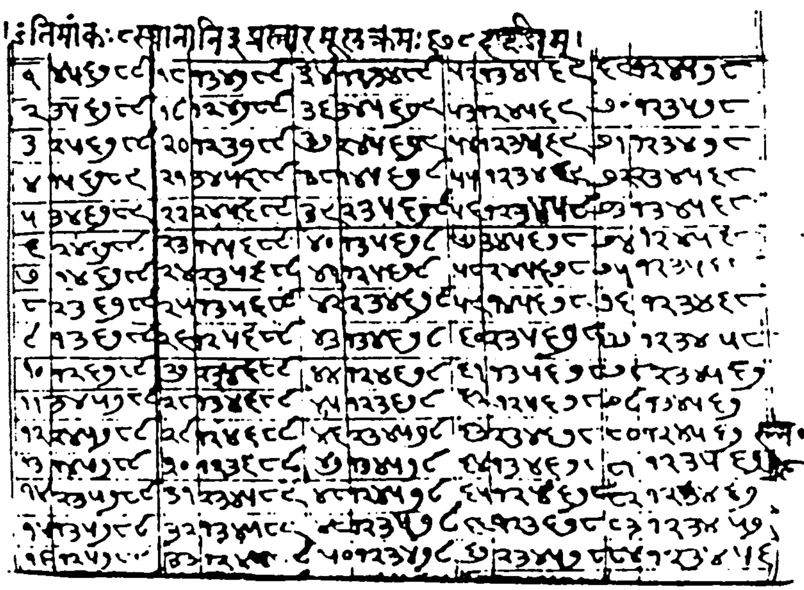
He divided his explanations and illustrations of these computations according to five different conditions, or 'patterns', determining the combinatorial nature of the problem, as follows:

1. n distinct digits in n places;
2. non-distinct digits in n places;
3. n places containing any digits from 1 up to a specified highest digit q;
4. any number of places containing any digits from 1 up to a specified highest digit q, where the sum of the digits in each variation is a given number s;
5. n places containing any n distinct digits from 1 up to a specified highest digit q.

Some of the *pratyayas* apply to all of the patterns, and others only to certain ones.



A Nepalese Sanskrit manuscript of Nārāyaṇa's *Gaṇitakaumudī* includes diagrams to illustrate the worked examples in his commentary: (a) the pattern-1 *prastāra*; (b) a pattern-3 *prastāra*, where the number of pieces is 6 and the highest digit is 2; (c) adding up the total number of variations in a pattern-4 *prastāra*, where the sum of the digits is 7 and the highest digit is 3.



A pattern-5 *prastāra*, where the number of places is 6 and the highest digit is 9, from a 1792 manuscript of Nārāyaṇa's *Gaṇitakaumudī*.

Nārāyaṇa's methods were designed for computational ease and modularity – that is, the sequences were used to determine the entries in the cells of a diagram, and the diagram entries and sequence elements were used together to solve problems in the calculation of the *pratyayas*. For example, the 'partial-*meru*' figure of size n (shown below for the case $n = 6$) is a half-grid whose top row contains 1 in its first cell followed by $n - 1$ zeros. Below that, as Nārāyaṇa described it, the columns all contain elements of the 'increasing sequence' $\{1, 2, 3, \ldots\}$ written from top to bottom and multiplied by 'the product of their own rows' (see [10, p. 328]) – that is, the cell in the ith row of the jth column (where $2 \le i \le j \le n$) contains the number $(i - 1) \times (j - 1)!$.

1	0	0	0	0	0
	1	2	6	24	120
		4	12	48	240
			18	72	360
				96	480
					600

This figure can then be used to answer different questions about the *pratyayas*, for variations with up to n places. For instance, if we wish to know the total number of variations produced by permuting the four distinct digits 2, 3, 6, and 1 (a pattern-1 case), we obtain it from what Nārāyaṇa [10, p. 328] called 'the sum of the numbers in the cells of the hypotenuse' for the first four columns: $1 + 1 + 4 + 18 = 24 = 4!$; equivalently, we can just multiply together the first four terms in the increasing sequence. The number of variations ending with each of the specified digits is $\frac{24}{4} = 6$, so the sum of the upper digits in all the variations is

$$(6 \times 6) + (6 \times 3) + (6 \times 2) + (6 \times 1) = 72.$$

This result is then multiplied by the four-digit 'sequence without intervals' 1111 to give $72 \times 1111 = 79\,992$ as the sum of all the digits in all the variations (see [10, p. 334]).

The use of 'markers' to designate certain cells in a diagram can be illustrated by the procedure prescribed by Nārāyaṇa for finding the *uddiṣṭa*, the serial number of a variation in the above example, as follows (see [10, p. 338]):

> Whatever is the number at the end of the indicated [variation], as far as that is from the final [number] of the base [order], just as far in the cell below of the partial Meru one should place a marker. There is omission of this [last number] in both the base order and the indicated [variation]. [Do this] again as long as there is such a number. The sum of the numbers which fall in the cells occupied by the markers is the measure of the variations at the indicated [variation] combined with the first number.

For instance, if the serial number of the four-digit variation 6231 is sought, we first compare it with the 'base order' variation 1236 with the digits appearing in ascending order. In the indicated variation 6231 the 'end' or rightmost digit 1 is fourth from the right in the base order. So in the rightmost column of the partial *meru* for $n = 4$, shown below, we mark the entry 18 (represented by italics) in its fourth row.

1	0	0	0
	1	*2*	6
		4	12
			18

Next, the omission of 1 from both variations results in the truncated forms 623 of the indicated variation and 236 of the base order. The new rightmost number 3 in the indicated variation is second from the right in the new form of the base order, so in the adjacent third column we mark the entry 2 in the second row. The same process tells us where to mark entries in the second and first columns. Then we add up all the marked entries to find the serial number of the indicated variation 6231: $18 + 2 + 1 + 1 = 22$.

We can see from the list of variations shown below that the result is in agreement with the standard *prastāra*, which requires us to permute the first two digits before changing the third, the first three digits before changing the fourth, and so on, until the last variation appears in reverse base order or 6321.

1.	1236	7.	1263	13.	1362	19.	2361
2.	2136	8.	2163	14.	3162	20.	3261
3.	1326	9.	1623	15.	1632	21.	2631
4.	3126	10.	6123	16.	6132	22.	6231
5.	2316	11.	2613	17.	3612	23.	3621
6.	3216	12.	6213	18.	6312	24.	6321

The *prastāra* with base order 1236.

Our final example of Nārāyaṇa's rules involves a somewhat more complicated case, where the sum s of the digits and the highest digit q are specified but the number of digits is not fixed. Such a set of variations is equivalent to the partitions of s that include any of the digits from 1 up to q. Nārāyaṇa [10, p. 359] noted:

> The number at the end of the sequence called additive is the measure of the variations. [The numbers taken] in reverse order from its penultimate number are the variations ending with 1 and so on.

For example, the additive sequence for $s = 7$ and $q = 3$ is the above-mentioned sequence $\{1, 1, 2, 4, 7, 13, 24, 44\}$, meaning that there are forty-four such numbers between 133 and 1111111. Of these numbers, twenty-four have 1 as their final digit, thirteen have 2, and seven have 3.

Furthermore, Nārāyaṇa employed a sequence mysteriously called 'the underworld' to glean more details about these variations (see [10, p. 323, p. 360]):

> One should write 0 and the number 1 below the sequence called additive. Then [the number] above the last [number] is added to the sum of the numbers in equal[ly many] places as the highest [digit], in reverse order. In this way one should write numbers in front of those in all the places ... The numbers produced from the sequence [called] the underworld for the same highest [digit] and sum are [taken] in reverse order. They are the occurrence [of] the digits 1 and so on with that sum.

That is, the 'underworld' sequence begins with the numbers 0 and 1 written under the first two numbers in the corresponding additive sequence for the given s and q, and its remaining terms are successively computed by adding up its last q terms and the number in the additive sequence above its last term. Thus, in the example for $s = 7$ and $q = 3$ shown below, we find the underworld sequence terms

$$0 + 1 + 2 + 2 = 5, \; 1 + 2 + 5 + 4 = 12, \; 2 + 5 + 12 + 7 = 26,$$

and so forth. Taken in reverse order, they tell us that the variations contain 118 occurrences of the digit 1, 56 occurrences of the digit 2, and 26 of the digit 3.

1	1	2	4	7	13	24	44
0	1	2	5	12	26	56	118

It remains unclear as to why mathematicians like Bhāskara II and Nārāyaṇa Paṇḍita cast so much of their combinatorics in terms of the 'net of digits' – that is, the arithmetic properties of variations interpreted as integers. Scholars in other *śāstras*, such as prosody and music, continued to develop the *pratyayas* of their disciplines (see [1, transl., p. 59]), but they do not seem to have reinterpreted them as special cases of the mathematicians' *aṅkapāśa*. Nor does *aṅkapāśa* itself seem to have profoundly impinged on other areas of mathematics. It may well have been valued primarily as an interesting research area in its own right, particularly for its unique combination of very simple operations with challenging methods and problems. As Nārāyaṇa observed (see [10, p. 320, p. 373]), paraphrasing a similar sentiment expressed earlier by Bhāskara:

> [even though,] in this mathematics called *aṅkapāśa* there is no multiplication, no division, no square and square root, no cube and its root, no operation for subtracting fractional parts, no miscellaneous [rules] for fractions, no rule of three or rule of five, no areas and volumes, no equation of unknowns and so on ....

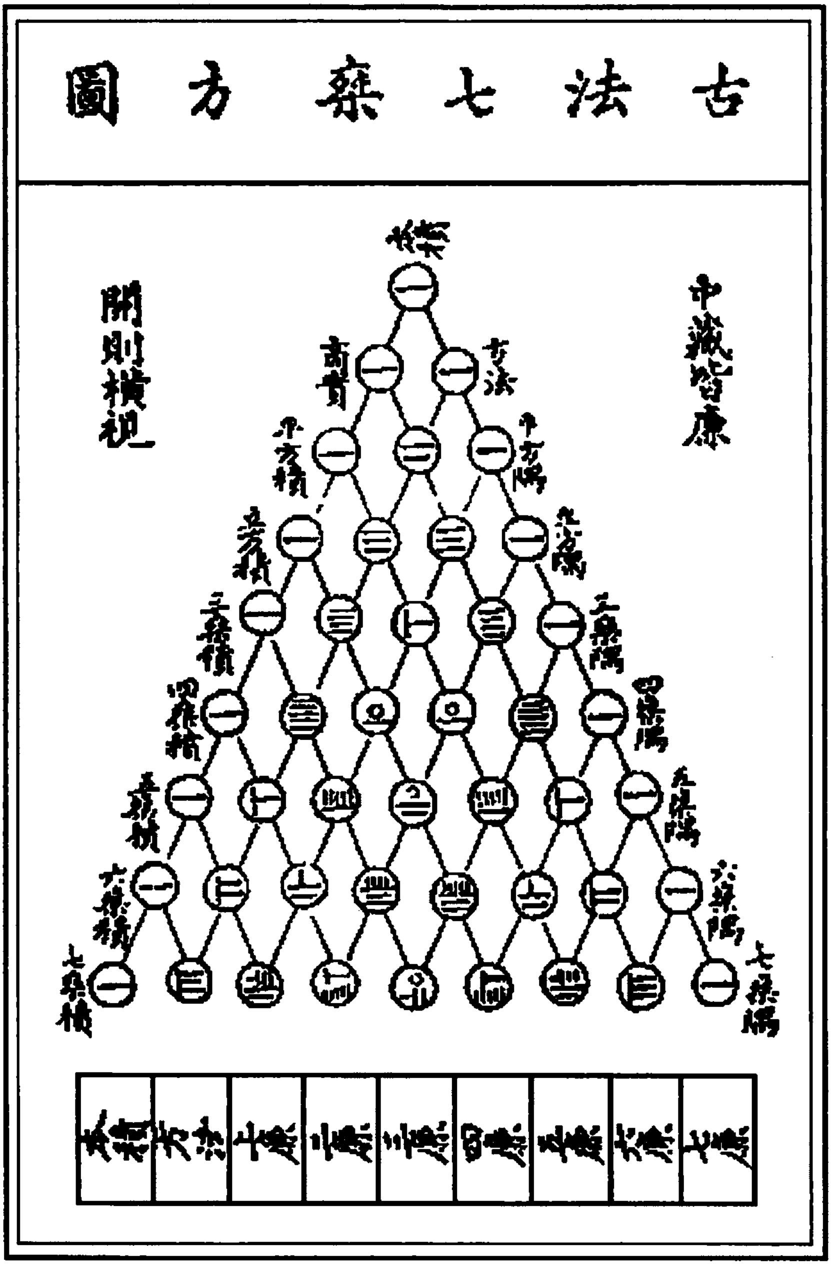
Nonetheless, in the net of digits 'those who are jealous, depraved, and poor mathematicians fall down'.

References

1. L. Alsdorf, Die pratyayas, *Z. Indologie Iranistik* 9 (1933) 97–153; transl. by S. R. Sarma, The pratyayas: Indian contribution to combinatorics, *Indian J. Hist. Sci.* 26 (1991), 17–61.
2. V. G. Āpaṭe, *Līlāvatī*, 2 vols. (Ānandāśrama Sanskrit Series 107), Ānandāśrama Press (1937).
3. S. Dvivedī, *Brāhmasphuṭasiddhānta* (*The Pandit* NS 23–4), Government Sanskrit College (1901–2).
4. T. Hayashi, Indo ni okeru junretu, kumiawase, rekkyo (Permutations, combinations, and enumerations in India), *Kagakushikenkyu* 131 (1979), 14–53 (in Japanese).
5. T. Hayashi, Varāhamihira's pandiagonal magic squares of the order four, *Historia Math.* 14 (1987), 159–66.
6. T. Hayashi, Combinatorics in Indian mathematics, *Encyclopaedia of the History of Science, Technology and Medicine in Non-Western Cultures* (ed. H. Selin), Kluwer (1997), 229–30.

7. T. Hayashi, Magic squares in Indian mathematics, *Encyclopaedia of the History of Science, Technology and Medicine in Non-Western Cultures* (ed. H. Selin), Kluwer (1997), 529–36.
8. N. A. Jairazbhoy, *Svaraprastāra* in north Indian classical music, *Bull. School of Oriental and African Studies* 24 (1961), 307–25.
9. D. Knuth, *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*, 3rd edn., Addison-Wesley, 1997.
10. T. Kusuba, *Combinatorics and Magic Squares in India*, PhD thesis, Brown University (1993).
11. K. Plofker, *Mathematics in India*, Princeton University Press (2009).
12. M. Rangacarya, *Ganita-Sara-Sangraha*, Government Press (1912).
13. E. Sachau, *Alberuni's India*, 2 vols, Munshiram Manoharlal (reprinted 1992).
14. P. Singh, The so-called Fibonacci numbers in ancient and medieval India, *Historia Math.* 12 (1985), 229–44.
15. P. Singh, The *Gaṇitakaumudī* of Nārāyaṇa Paṇḍita, Chapter XIII, *Gaṇita Bhāratī* 23 (2001), 18–82.
16. R. Sridharan, Sanskrit prosody, Pingala *sūtras* and binary arithmetic, *Contributions to the History of Indian Mathematics* (ed. G. Emch *et al.*), Hindustan Book Agency (2005), 33–62.

This page intentionally left blank



The arithmetical triangle from Zhu Shijie's *Siyuan yujian* of 1303.

CHAPTER 2

China

ANDREA BRÉARD

Combinatorial practices in China go back to high antiquity, when divinatory techniques relied on configurations of broken and unbroken lines. The *Yijing* or *I Ching* (Book of Change), compiled under the Zhou dynasty, has transmitted these practices up to the present time and has been a widely commented upon and read source. But combinatorial practices in China were not limited to divination and magic squares: a large number of early sources also described games such as Go and chess, and games with cards, dominoes, and dice, that show a combinatorial interest from a more mathematical point of view. The earliest source that systematically discusses permutations and combinations is an 18th-century manuscript. Although mathematics had by then been introduced from Europe, the manuscript is clearly based on traditional mathematical concepts and algorithmic modes. In this chapter we show how early combinatorial practices provided a framework for later mathematical developments in imperial China.

Combinatorial practices

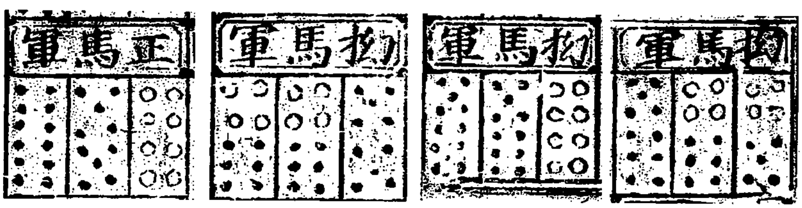
We begin with a brief look at three areas in which combinatorial practices occurred: divination, games, and magic squares.

Divination

An early interest in combining and arranging two distinct elements as n-tuples can be found in one of China's Confucian classics, and goes back to as early as the end of the Zhou dynasty (*c.* 3rd century BC). As mentioned in Donald Knuth's introductory chapter, the *Book of Change* gives – in a specific sequence that changed over time – interpretations of the sixty-four hexagrams, diagrams of six stacked broken or unbroken lines. Leibniz's prominent arithmetical reading of these figures as a binary system certainly was a misinterpretation, though by no means the sole occurrence of linking these diagrams used for divination with mathematical content. The following sections show how Chinese authors later used them as a model to discuss combinatorial issues from a theoretical point of view (see [1]).

Gaming

During the Song dynasty (960–1279), gaming emerged as another field of combinatorial practice in relation to mathematical writing. Shen Gua (1031–95), a polymath and state official, explicitly discussed the possible configurations in the game of Go, with a grid of 19×19 lines, where each position could be empty, or could contain a black stone or a white stone. Furthermore, late-16th-century texts, generally referred to by historians as *Riyong leishu* (Encyclopedias for Daily Use), describe the game of 'ivory tiles' (*yapai*).

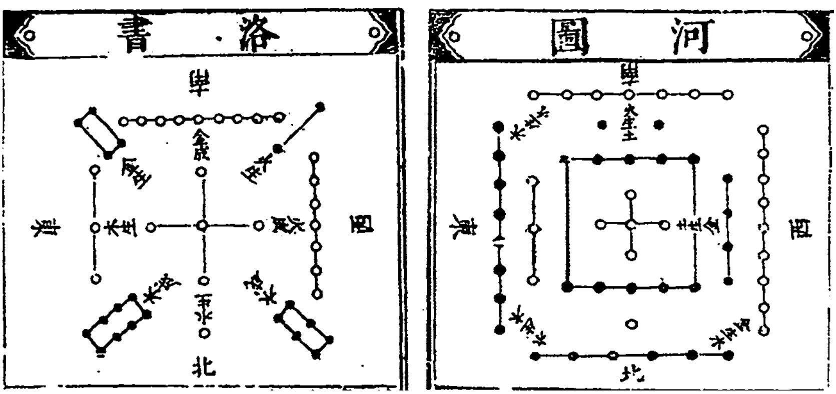


A late-16th-century manual showing the possible permutations of pips on three 'ivory tiles', each depicting two of the numbers 4, 5, and 6.

The origins of these domino-like tiles, played in a card-like fashion, date back to the Song dynasty Xuanhe era (1082–1135). The printed pattern sheets attest to a certain combinatorial activity, that of enumerating the possible permutations of the number of pips on a combination of three tiles, as shown above. The three doubles shown on the left were called the 'regular cavalry', whereas all other permutations were referred to as 'irregular cavalries'. It is possible that winning schemes in the game of ivory tiles were based on combinatorial considerations, but there is certainly no evidence of the existence of a concept of probability (see [2]).

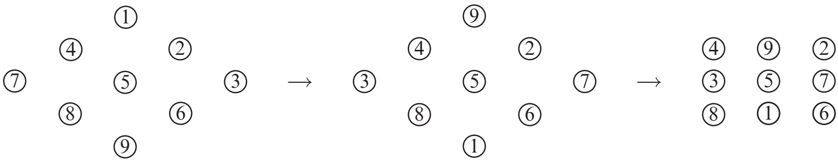
Magic squares

Magic squares have also been an often-cited reference for combinatorial theories in early China. But, as Cammann [3] has pointed out, only two diagrams, the *Hetu* (Yellow River Chart) and the *Luoshu* (Luo River Writing), which are legendarily ascribed to two semi-divine figures from the 3rd and 2nd millenia BC, appear in the surviving mathematical and other texts before the Song dynasty. In the following figure, 'knotted chord' configurations are used to represent numbers. The Yellow River Chart on the right shows an arrangement of the numbers 1 to 10, whereas only the Luo River Writing on the left is a true 3×3 magic square of the numbers 1 to 9. Both diagrams were deeply rooted in a correlative system of thought, relating odd and even numbers to *yin* and *yang*, and their positions to the five directions (the four cardinal directions and the centre), five elements, or four seasons.



The *Luoshu* (left) and *Hetu* (right) diagrams, from Cheng Dawei's *Unified Lineage of Mathematical Methods* (1592).

Yang Hui (*c.*1238–98), in his *Xugu zhaiqi suanfa* (Continuation of Ancient Mathematical Methods for Elucidating the Strange Properties of Numbers) (1275), gave a method for constructing a magic square of order 3 (resulting precisely in the *Luoshu*) and two of order 4. For order 3, Yang Hui started from a diagonal arrangement (see left figure below), and then prescribed interchanging top with bottom, and right with left. ‘Making the four corners stick out’ results in the arrangement of the *Luoshu* (see right figure below and left figure above).



Construction of the *Luoshu* diagram in Yang Hui (1275).

In the case of a 4 × 4 magic square, an analogous method was presented without justification to produce the example shown in the right figure below. Here, Yang Hui started from an arrangement of the consecutive numbers 1 to 16 in the four columns of a square array and again proceeded by interchanging the corners – first the corners of the outer square, then the corners of the inner square – and in the end any row of the diagram adds up to 34, whether vertical, horizontal, or diagonal, as shown below.

13	9	5	1
14	10	6	2
15	11	7	3
16	12	8	4

→

4	9	5	16
14	7	11	2
15	6	10	3
1	12	8	13

Furthermore, Yang Hui showed two particular examples each of magic squares of orders 5 to 8, and one example each of orders 9 and 10. One magic square of order 5 and one of order 7 are ‘bordered’: one or more of the inner magic squares are magic squares themselves (see the figure below). Such examples can be found in earlier sources from the Islamic world, but, as Yang Hui did not even mention this property of his 5 × 5 and 7 × 7 squares, it is unclear whether such particular magic squares were transmitted to China or discovered independently.

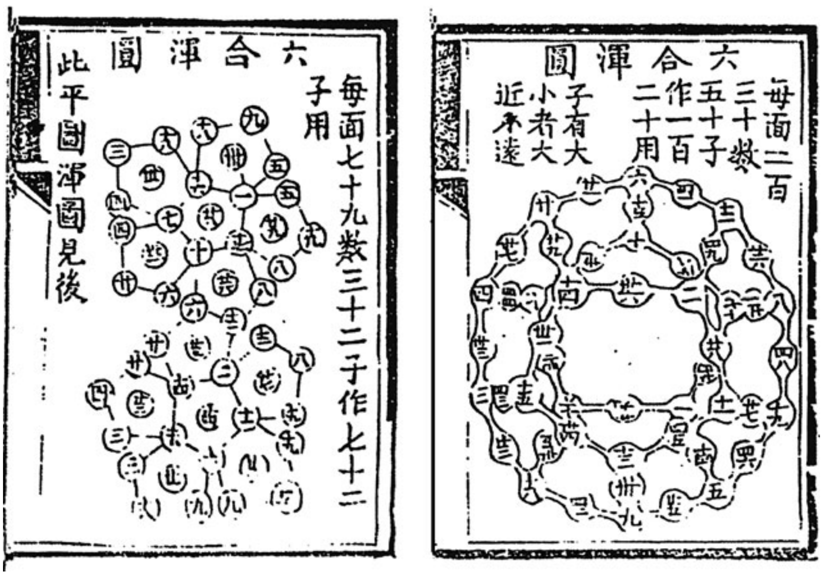
1	23	16	4	21
15	14	7	18	11
24	17	13	9	2
20	8	19	12	6
5	3	10	22	25

46	8	16	20	29	7	49
3	40	35	36	18	41	2
44	12	33	23	19	38	6
28	26	11	25	39	24	22
5	37	31	27	17	13	45
48	9	15	14	32	10	47
1	43	34	30	21	42	4

Two bordered magic squares given by Yang Hui.

A remarkably original, but little known, contribution to magic squares occurred in the late 1800s in China. Bao Qishou found ‘magic perimeter solutions’ for the five Platonic solids: the cube, the tetrahedron, the octahedron, the icosahedron, and the dodecahedron. The author seems to have been interested in generalizing the problem of assigning a number to each vertex of a solid in such a way that consecutive numbers on the perimeters of the structure all sum to the same constant. For example, Bao labelled the vertices of a cube with the numbers 1 to 8 so that the sum around each face is 18, and also labelled the edges with 1 to 12 so that each sum is 26; then, by combining these two labellings (adding 8 to the edge labels), he achieved a perimeter labelling, using 1 to 20, where each sum is 76. But Bao did not reveal his secret of how precisely one might proceed, and rather defied the reader to solve the problem [4]:

Further diagrams can be constructed by the previous methods. It becomes easier to make variations when more and more numbers are used. However, it is pointless to give these illustrations when all methods have been exhaustively demonstrated.



Two ‘magic’ labellings of the vertices of a dodecahedron by Bao Qishou.

In his essay *Zengbu suanfa hunyuantu* (Additional Mathematical Methods for Solid and Spherical Figures), Bao used a variety of diagrams, unfoldings, or three-dimensional illustrations to depict his perimeter magic figures. Above we see the particularly interesting case of the dodecahedron. On the left, we see a planar representation with the vertices consecutively labelled from 1 to 20, with the twelve pentagonal faces summing to the twelve numbers from 47 to 58. This figure is not a magic dodecahedron, but is still a clever variation, since it is impossible to have constant face sums with consecutive-vertex labelling. The three-dimensional illustration on the right in the figure above is another variant, in which the vertices of the dodecahedron have the same labelling with 1 to 20 as on the left, and the numbers 21 to 50 for the edges. Thus, Bao obtained a fully 'face-magic' dodecahedron in which each face sum is 230.

Chen Houyao's manuscript

Among the now extant Chinese mathematical writings, there is only one manuscript essay devoted to combinatorics: *Cuozong fayi* (The Meaning of Methods for Alternation and Combination) by Chen Houyao (1648–1722). It dealt systematically with problems of permutations and combinations in the case of divination with trigrams, the formation of hexagrams or names with several characters, and combinations of the ten heavenly stems (*tiangan*) and the twelve earthly branches (*dizhi*) to form the astronomical sexagesimal cycles. Games of chance, such as dice throwing and card games, equally serve as a vehicle for discussing algorithms for calculating combinations, with or without repetition. In the foreword to his treatise, Chen Houyao underlined the originality of his contribution to the mathematical tradition in China [5]. Referring to the classic *Nine Chapters of Mathematical Procedures* [6], he observed:

> The *Nine Chapters* have entirely provided all [mathematical] methods, but they lack of any type of method for alternations and combinations.

Unfortunately we do not know whether the author based his solution methods on knowledge circulating among early Qing dynasty mathematical networks, or whether he learned about combinatorics from the teachings he received from the Kangxi Emperor, who in turn had been instructed by the French Jesuit mathematicians sent by Louis XIV, 'Les mathématiciens du Roi'.

Although singular in its appearance, the existence of Chen Houyao's short but well-constructed collection of problems and solution procedures is nevertheless an indication that other sources with combinatorial problems may have circulated during the late imperial period. Similar phenomena of selective textual transmission in China can be observed for the development of other mathematical concepts and ideas. Another reason to believe that combinatorics was a wider field of mathematical enquiry is the fact that Chen Houyao referred explicitly to his predecessors when he expressed his intention to improve the efficiency of their algorithms.

In the case of hexagrams (configurations made up of six lines, each broken or unbroken, as discussed in Knuth's introductory chapter), he underlined that the calculation of all possible combinations with repetition can be obtained either by successive multiplication of the two possibilities:

$$\begin{aligned}
\text{number of configurations consisting of 2 lines} &= 2 \times 2 = 4;\\
\text{number of configurations consisting of 3 lines} &= 4 \times 2 = 8;\\
&\vdots\\
\text{number of configurations consisting of 6 lines} &= 32 \times 2 = 64;
\end{aligned}$$

or, in a much simpler way, by squaring the eight possibilities for obtaining a trigram (a configuration made up of three lines) (see [7]):

If one multiplies by itself the thus obtained number [the number 8 for trigrams], one economizes half of the multiplications.

In other problems in which Chen Houyao suggested two algorithms for finding the solution, he not only named one of them explicitly as the 'original method', but also indicated the detailed sequence of operations to follow. For one problem, the equivalence (in modern mathematical terms) between Chen's method,

$$\begin{aligned}
C(30,9) &= \frac{30 \times 29 \times 28 \times 27 \times 26 \times 25 \times 24 \times 23 \times 22}{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2}\\
&= \frac{5\,191\,778\,592\,000}{362\,880} = 14\,307\,150,
\end{aligned}$$

and the 'original method',

$$C(30,9) = (\ldots(((22 \times 23) \div 2) \times 24) \div 3) \times \cdots \times 30) \div 9,$$

is illustrated in the following example [8]:

Let us suppose one has thirty playing cards. All cards have a [different] form. Each hand consists of nine cards. How many combinations can one obtain by drawing one hand? It says: 14 307 150 hands.

The method says: One takes thirty cards as the dividend. Furthermore one subtracts one from thirty. This makes twenty-nine, one multiplies this [the dividend of 30] with it. One then again subtracts one, and multiplies this [the product of 30 and 29] with 28 . . . One multiplies this [the product of 30, 29, 28, 27, 26, 25, 24, and 23] with 22. Nine cards make up one hand. Because one subtracts nine layers, one has to multiply eight times to obtain the dividend. Altogether one obtains by multiplications 5 191 778 592 000 as the dividend. Now one shall also successively subtract each hand of the nine cards. One multiplies this [the nine cards] with 8, further one multiplies this with 7 . . . Further one multiplies this [the product of 9, 8, 7, 6, 5, 4, and 3] with 2. With one [multiplied] this remains unchanged, therefore one does not multiply. Altogether one obtains by these multiplications 362 880 as divisor. By this one divides the above dividend, one obtains 14 307 150 hands. The explanation (*jie* [9]) says: Here the situation is as in the previous example concerning the combination of eight personal names. Yet although one does not have repetitions in combinations of names, one can still invert their ordering. But here each hand has nine cards, each card has a different colour, and there are neither repetitions nor permutations of their ordering. This is the reason why the methods which are used are different. At first, the calculation of 5 191 778 592 000, which one obtains through successive subtraction and multiplication of the thirty cards, equals the previous method concerning the combination of personal names. Names have no above or below, no inversion of ordering. That is why one further has to eliminate the equivalent ones obtained by changing the ordering. Thus one divides this [the dividend 5 191 778 592 000] by the divisor, the successively subtracted and multiplied nine cards. One then obtains the real number.

The original method multiplies and divides alternately one at a time. Its principle is not easy to grasp and the method is very confusing and clumsy. It does not equal the efficiency of common multiplication and common division in this method here.

This problem shows that in China games of chance did provide a framework for the mathematical treatment of possible outcomes. But it is doubtful that the conceptual step that relates the number of favourable events and the total number of possible events, which in Europe laid the foundations of a mathematical theory of probability, was ever taken in China. Neither Chen's method, nor later algorithms to calculate combinations and permutations, seemed to inspire further theoretical considerations of chance, at least judged on the mathematical sources preserved. In China probability theory, now an important part of mathematical statistics and a field that relies on combinatorial tools, seems to have been a mathematical area entirely imported from the West.

The arithmetical triangle

An even earlier prominent candidate for a possible emergence of combinatorial considerations in the Chinese mathematical tradition might be what is known as the 'Pascal triangle' or the 'arithmetical triangle'. It first appeared in China in a chapter on algorithms for root extraction, in Yang Hui's *Xiangjie jiu zhang suanfa* (Detailed Explanations of The Nine Chapters on Mathematical Methods), completed in 1261, but it must have been circulating a century earlier. In the West, Pascal also applied his arithmetical triangle to the theory of combinations, the powers of binomial quantities, and the *problem of points* – the division of stakes in an interrupted game of chance (see [10], [11], and Chapter 6).

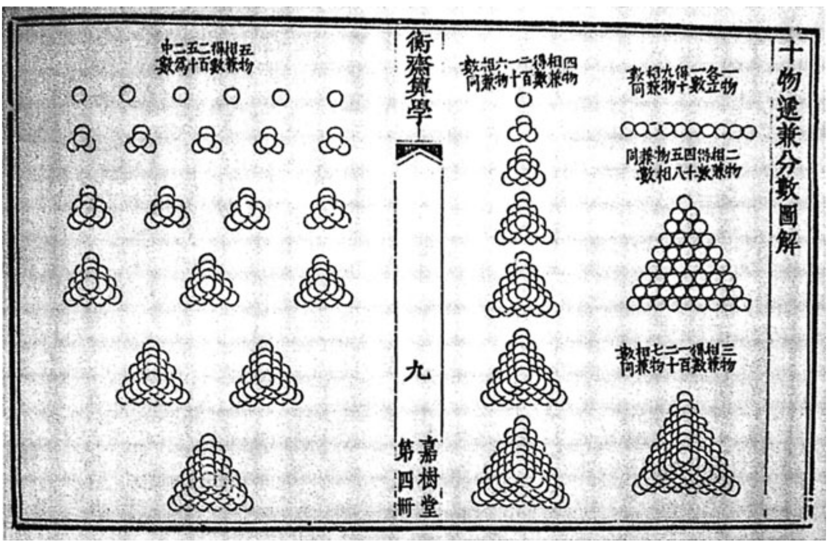
In China, we know only of its important use in the context of interpolation techniques, solutions of polynomial equations, and the construction of finite arithmetical series – in particular, through an early-14th-century treatise, the *Siyuan yujian* (Jade Mirror of Four Elements, 1303) by Zhu Shijie, who placed the diagram at the beginning of his book. Although Zhu did not give a combinatorial interpretation of the diagram, his interest in combinatorial practices played out on another level: the terminology used for the finite series was constructed from combinations of binomial expressions in the linguistic sense. A mathematical meaning was attached to each prefix or suffix, reflecting the factorization of each term in the arithmetical series. Inverse problems (such as how many terms form a given sum) were solved by interpolation techniques, using the coefficients from the triangle (see [12, Ch. 4.2] and [13]).

Later, Wang Lai (1768–1813) did relate finite arithmetical series to combinatorial problems in an essay *Dijian shuli* (The Mathematical Principles of Sequential Combinations) [14]. Without referring to the arithmetical triangle, he illustrated his subject matter with the example of ten objects from which sequentially one, two, three, four, or five objects are drawn. He was probably unaware of Chen Houyao's manuscript, since he claimed the same priority to the field of combinatorics [16]:

Procedures of sequential combinations had not been discovered in ancient times. Now that I have decided to investigate them, it is thus appropriate to explain the object of inquiry

first. Let us suppose one has all kinds of objects. Starting off from one object, of which each establishes one configuration, and going up to all the objects taken together, they form altogether one configuration (*shu*). In between lie sequentially:
two objects connected to each other form one configuration, and we shall discuss how many configurations this can make through exchanging and permuting (*jiao cuo*);
three objects connected to each other form one configuration, and we shall discuss how many configurations this can make through exchanging and permuting;
four objects, five objects, up to arbitrarily many objects, they all entirely follow that which is the so-called *procedure of sequential combinations*.

The possible outcomes of drawing k objects from a set of n objects correspond to sums of higher-order series, which Zhu Shijie had already calculated in his *Siyuan yujian*, but without explicitly referring to problems of combination. Here, for the first time in the transmitted Chinese mathematical tradition, Wang Lai linked combinations to figurate numbers. He gave drawings for $C(10, k)$, illustrating the sum of finite series with surfaces and piles of unit pebbles, as shown below, for $k = 1$ to 5, from right to left. He also remarked on the symmetrical relation $C(10, k) = C(10, 10 - k)$, which explains why he did not illustrate the cases $C(10, k)$ for $k = 6$ to 10.



Wang Lai's *Dijian shuli*.

When Chen calculated the total number of pebbles arranged in triangular or pyramidal shapes – the so-called 'triangular piles' (*sanjiao dui*) – he used Zhu

Shijie's procedures for calculating the sums of finite arithmetical series of higher order:

$$\begin{aligned}
C(10,1) &= C(10,9) = 1 + 1 + \cdots + 1 = 10;\\
C(10,2) &= C(10,8) = 1 + 2 + 3 + \cdots + 9 = \frac{9 \times 10}{2} = 45;\\
C(10,3) &= C(10,7) = 1 + 3 + 6 + 10 + \cdots + 36 = \frac{8 \times 9 \times 10}{2 \times 3} = 120;\\
C(10,4) &= C(10,6) = 1 + 4 + 10 + 20 + \cdots + 84 = \frac{7 \times 8 \times 9 \times 10}{2 \times 3 \times 4} = 210;\\
C(10,5) &= 1 + 5 + 15 + 35 + 70 + 126 = \frac{6 \times 7 \times 8 \times 9 \times 10}{2 \times 3 \times 4 \times 5} = 252.
\end{aligned}$$

The illustrations of $C(10, k)$ shown above (where $k = 1$ to 5) start in the right-hand column and suggest the patterns of formation of every term of the series, beginning with $C(10, 1) = 1 + 1 + \cdots + 1 = 10$. These illustrations stemmed from the Yuan dynasty tradition of considering piles of discrete objects in different geometrical shapes. The series $C(10, 2) = 1 + 2 + \cdots + 9$ thus became a triangle in which pebbles were piled up in rows with one to nine pebbles in each successive row. The next sum, for $C(10, 3)$, was then a regular pyramid, where each layer was composed of one such triangle with (from top to bottom) $1, 3, 6, \ldots, 36$ elements. For $C(10, 4)$, Wang showed seven pyramids in the left-hand column of the right-hand page:

$$C(10,4) = 1 + (1+3) + (1+3+6) + \cdots + (1+3+6+10+15+21+28).$$

Finally, on the left-hand page $C(10, 5)$ was thought of as twenty-one pyramids that can be grouped in two ways. A horizontal reading of the drawing gives the terms

$$C(10,5) = (6 \times 1) + (5 \times 4) + (4 \times 10) + (3 \times 20) + (2 \times 35) + (1 \times 56),$$

whereas a diagonal reading from right to left produces different terms for the same sum:

$$\begin{aligned}
C(10,5) = 1 &+ (1+4) + (1+4+10) + (1+4+10+20)\\
&+ (1+4+10+20+35) + (1+4+10+20+35+56).
\end{aligned}$$

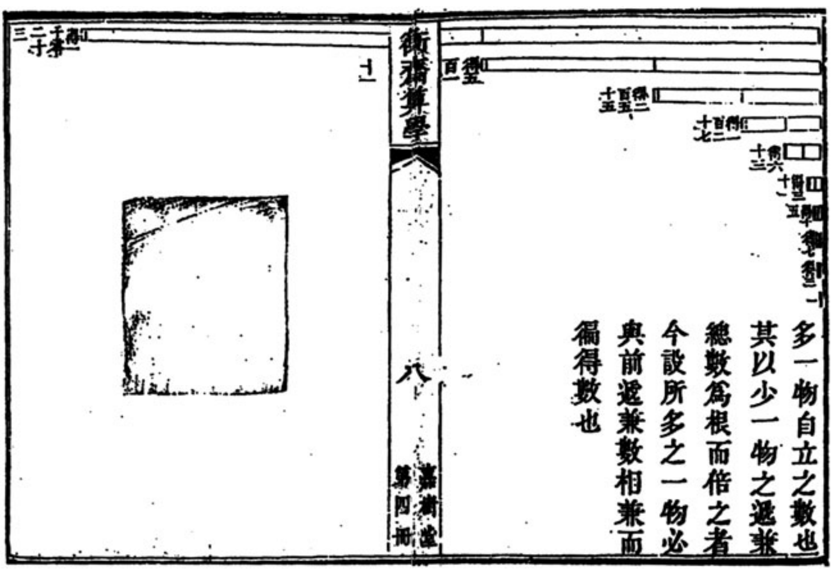
The author then illustrated a general method for calculating the total sum of combinations

$$s_n = C(n,1) + C(n,2) + \cdots + C(n,n).$$

The indicated algorithm corresponds in modern mathematical terms to an iterative procedure: successively double the 'root' (the preceding result) and add unity. Given a set of n objects, and starting with $s_1 = 1$, Wang prescribed iterating the following operation for $k = 2, 3, \ldots, n$:

$$s_k = 2s_{k-1} + 1.$$

The corresponding figure (see below) depicts these $n - 1$ iterations for $n = 10$, successively doubling a horizontal bar in length and extending it by a unitary amount. One thus obtains $s_{10} = 1023$.



Some historians [17] have claimed that Wang Lai recognized here the remarkable identity

$$C(n,1) + C(n,2) + \cdots + C(n,n) = 2^n - 1,$$

but no explicit mention of the fact that $1023 = 2^{10} - 1$, nor of any kind of generalization, can be found in his text.

As an application of his procedure for calculating s_n, Wang Lai stated a related mathematical problem. It stems from the earliest instance of combinatorial practices, mentioned at the beginning of this chapter: divination with broken and unbroken lines. In Wang's example, a shaman performing yarrow stalks divination (*shigua*) produces a hexagram, a configuration made up of six lines (*liu yao*). Wang was interested in the total number of possible configurations made up of 1 to 6 lines that one can produce from this hexagram; this corresponds to finding the sum

$$C(6,1) + C(6,2) + \cdots + C(6,6).$$

He calculated his result in two distinct ways. The first method proceeds by doubling successively the result for the maximum number of lines in such a configuration, and then adding 1. Wang Lai remarked that five iterations give the total number of possible configurations. Thus, in five steps he calculated the final result:

$$(2 \times 1) + 1 = 3; \quad (2 \times 3) + 1 = 7; \quad (2 \times 7) + 1 = 15;$$
$$(2 \times 15) + 1 = 31; \quad (2 \times 31) + 1 = 63;$$

or, when transcribed into a condensed formula,

$$63 = 2 \times \Big(2 \times \Big(2 \times \big(2 \times (2 \times 1 + 1) + 1\big) + 1\Big) + 1\Big) + 1.$$

Alternatively, Wang Lai could have calculated the total of 63 configurations by adding up the various possibilities for the configurations in which $1, 2, \ldots, 6$ lines were used. He indeed calculated each of these by using the older procedures for 'triangular piles' (procedures for calculating the sums of triangular numbers, as already found in Zhu Shijie's text of 1303). Again, as Chen Houyao had done earlier, Wang Lai remarked on the symmetry $C(n,k) = C(n, n-k)$,

$$\begin{aligned} C(6,1) &= C(6,5) = 6, \\ C(6,2) &= C(6,4) = 5 \times \tfrac{1}{2}(5+1) = 15, \\ C(6,3) &= 4 \times (4+1) \times \tfrac{1}{6}(4+2) = 20, \\ C(6,6) &= 1, \end{aligned}$$

but did not explicitly calculate their sum,

$$C(6,1) + C(6,2) + \cdots + C(6,6) = 6 + 15 + 20 + 15 + 6 + 1 = 63.$$

As mentioned earlier, Wang Lai did not connect his calculations with the arithmetical triangle. Its seventh line would contain precisely the numbers $1, 6, 15, 20, 15, 6$, and 1 ($C(n, 6)$, for $n = 0, 1, \ldots, 6$), and their sum equals 2^6. But Wang did not refer here to the corresponding values in the triangle, and no explicit mention of the fact that $63 = 2^6 - 1$, nor of any kind of generalization to $s_n = 2^n - 1$, can be found in his text. What Wang gave in the end was a general procedure for the sum of higher-order 'triangular piles' (finite arithmetic series):

$$\frac{(n+1) \times (n+2) \times \cdots \times (n+k)}{1 \times 2 \times \cdots \times k} = C(n+k, k).$$

As an example, he explicitly formulated the procedure and performed numerical calculations to determine the sum of the so-called 'fourth-order triangular pile' (*si cheng sanjiao dui*) [18], with 5 as the particular 'base number' (*genshu*):

$$C(4+5, 5) = \frac{5 \times (5+1) \times (5+2) \times (5+3) \times (5+4)}{1 \times 2 \times 3 \times 4 \times 5} = 126.$$

Li Shanlan

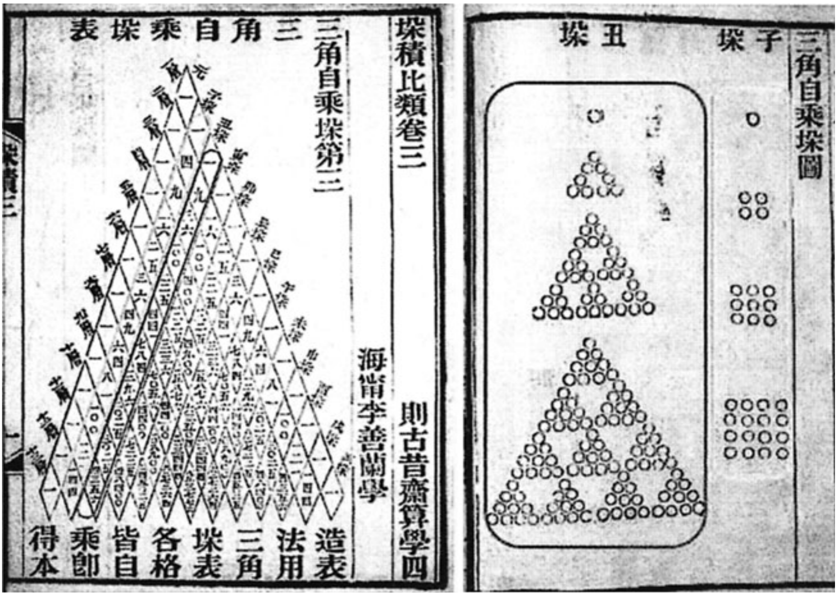
In 1867 number-theoretic relationships in arithmetical triangles were examined more systematically in China by Li Shanlan (1811–82) in his *Duoji bilei* (Analogical Categories of Discrete Accumulations) [19]. The most important result this led to is the famous 'Li Renshu identity' (see [20]), now written as

$$\sum_{k=0}^{n} C(n, k)^2 \, C(m + 2n - k, 2n) = C(m + n, n)^2.$$

Naturally, Li Shanlan did not present his identity in this way. In fact, he did not use any algebraic formalism to develop his formula at all, but expressed it entirely in traditional algorithmic and rhetorical language. His book is a deductive construction of generalized arithmetical triangles, starting from the Pascal triangle. The sum of the cubes of natural numbers, for example, was deduced as a sum of multiples of certain diagonals of the Pascal triangle [21]:

$$\sum_{k=1}^{n} k^3 = \sum_{k=1}^{n} C(k+2, 3) + 4 \times \sum_{k=1}^{n} C(k+1, 3) + \sum_{k=1}^{n} C(k, 3).$$

In the case of the 'Li Renshu identity', Li gave a generalized arithmetical triangle showing the square binomial coefficients $C(k+n,n)^2$. Thus the second diagonal from the left in the figure below shows, for $k = 0, 1, \ldots, 11$, the values (from top to bottom) of $C(k+1,1)^2$, and the third diagonal from the left shows, for $k = 0, 1, \ldots, 10$, the values of $C(k+2,2)^2$.



Li Shanlan's *Duoji bilei*.

Like Wang Lai, Li Shanlan added geometrical representations of the cells of his generalized arithmetical triangles in the form of figurate numbers. The figure on the right above shows the quadratic numbers – the coefficients $C(k+n,n)^2$ for the values $n = 1$ and $k = 0, 1, 2, 3$ on the right, and the corresponding coefficients $C(k+n,n)^2$ for the values $n = 2$ and $k = 0, 1, 2, 3$ as triangular patterns of triangles on the left. The Li Renshu identity itself was given implicitly in the statement on the construction of the cells of the generalized triangle in the figure as the sum of cells in previously constructed triangles.

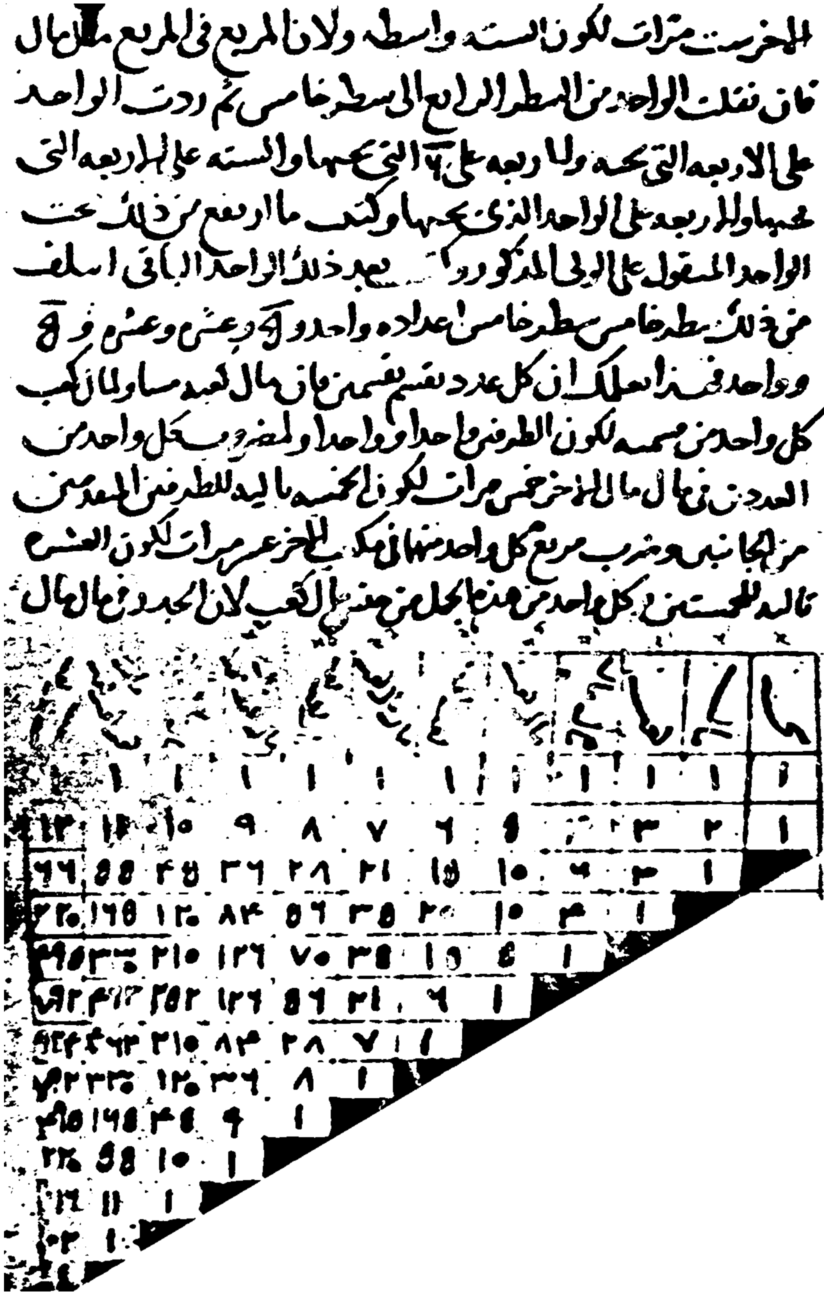
But interpreting Li Shanlan's writings as contributions to combinatorics is problematic (see [22]). His work, as viewed from within the Chinese mathematical tradition, is rather situated in a specific mathematical domain that developed as early as the Han dynasty. In particular, it was concerned with the discretization of continuous solids into a finite number of elements. Gradually abstracted from a specific geometrical context, this ultimately led Zhu Shijie in 1303 to work out a systematic set of procedures for the summation of finite series in relation to the arithmetical triangle. Several Qing dynasty authors, such as Li Shanlan, pursued research along these lines, but the field evolved without any combinatorial interpretation (see [12], Ch. 5).

From a present-day mathematical perspective, their results, taken out of context and transcribed into modern algebraic formulas, can indeed be seen as important combinatorial contributions. We have tried to show, by taking into consideration earlier mathematical traditions in China, that Chen Houyao and Wang Lai's texts turn out to be the only two that conveyed a genuine combinatorial meaning. Chen's manuscript is particularly original in that it presented new mathematical problems on games of chance and divination, two ancient and widely diverse fields of combinatorial practice in China. In Wang Lai's essay, the canonical hexagrams were established as the sole paradigmatic combinatorial model for the discussion of algorithms from a rational mathematical point of view.

Notes and references

1. A more detailed discussion of Leibniz's reading of the hexagrams, and the role that they played in the development of mathematical reflections, can be found in A. Bréard, Leibniz und Tschina – ein Beitrag zur Geschichte der Kombinatorik?, *Kosmos und Zahl; Beiträge zur Mathematik- und Astronomiegeschichte, zu Alexander von Humboldt und Leibniz* (ed. H. Hecht, I. Schwarz, H. Siebert, and R. Werther), Vol. 58 of *Boethius*, Steiner Verlag (2008), 59–70.
2. A more detailed description of the game and the texts that describe it can be found in A. Bréard, Knowledge and practice of mathematics in Late Ming daily life encyclopedias, *Looking at it from Asia: The Processes that Shaped the Sources of History of Science* (ed. F. Bretelle-Establet), Vol. 265 of *Boston Studies in the Philosophy of Science*, Springer (2010), 305–29, and A. Bréard, Usages et destins des savoirs mathématiques dans les *Encyclopédies aux dix mille trésors* des Ming, *Pratiques Lettrées au Japon et en Chine XVII^e–XIX^e Siècles* (ed. A. Horiuchi), Vol. 5 of *Études Japonaises, Les Indes savantes*, Paris (2010), 103–23.
3. S. Cammann, The evolution of magic squares in China, *J. Amer. Oriental Soc.* 80(2) (1960), 116–24.
4. Cited from p. 218 of Guo Shuchun *et al.* (ed.), *Zhongguo kexue jishu dianji tonghui. Shuxue juan*, 5 vols., Henan jiaoyu chubanshe, Zhengzhou (1993).
5. Translated from p. 685 of Chen Houyao, *Cuozong fayi* (The Meaning of Methods for Alternation and Combination); reprinted in [4], Vol. 4, 685–8.
6. Reference to *Jiu zhang suan shu* (Nine Chapters on Mathematical Procedures), the foundational and canonical work of mathematics in ancient China, compiled approximately during the 1st century AD.
7. Translated from [5], p. 685.

8. Translated from [5], p. 687.
9. Here, Chen Houyao used a graphic variation of the character for *jie*.
10. This was depicted in the *Traité du Triangle Arithmétique* 'apparently printed in 1654 (though circulated in 1665)'; see p. 9 of L. J. Daston, *Classical Probability in the Enlightenment*, Princeton University Press (1988).
11. Isaac Todhunter, as well as many other historians of probability theory, considered the problem of points (see Chapter 6), which prompted the seminal correspondence between Pascal and Fermat in 1654, as the beginning of the theory of probability. See Chs. 2 and 9 in I. Todhunter, *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Cambridge University Press (1865) (reprinted by Chelsea, 1965).
12. A. Bréard, *Re-Kreation eines mathematischen Konzeptes im chinesischen Diskurs: Reihen vom 1. bis zum 19. Jahrhundert*, Vol. 42 of *Boethius*, Steiner Verlag (1999).
13. A. Bréard, The reading of Zhu Shijie, *Current Perspectives in the History of Science in East Asia* (ed. S. K. Yung and F. Bray), Seoul National University Press (1999), 291–306.
14. Included as *juan* 4 of his collected writings, *Hengzhai suanxue* (Hengzhai's Mathematics) (reprinted in [4], Vol. 4, pp. 1512–16). For a discussion of its mathematical content, see [15, pp. 52–64].
15. Li Zhaohua, *Gusuan jinlun* (Ancient Mathematics – Contemporary Discussions), Tianjin kexue jishu chubanshe, Tianjin (1999).
16. Translated from [4], Vol. 4, pp. 1512–13, of Wang Lai, *Dijian shuli* (Mathematical Principles of Sequential Combinations), *Hengzhai suanxue* (Hengzhai's Mathematics), Jiashutang, China (1854), 6b–12b (reprinted in [4], Vol. 4, pp. 1512–16).
17. See, for example, Liu Dun's introduction to [16, p. 1479] or Li Zhaohua, Wang Lai 'Dijian shuli', 'Sanliang suanjing' lüelun (A short discussion of Wang Lai's 'Mathematical Principles of Sequential Combinations' and 'Mathematical Classic of Two and Three'), in Wu Wenjun (ed.), *Zhongguo shuxueshi lunwenji* (China Historical Materials of Science and Technology), Vol. 2 (1986), 65–78.
18. Literally, the Chinese term rendered here as 'order' can be translated as 'multiplication' since, as Chen pointed out himself, this number corresponds to the number of multiplications to be performed to calculate both the dividend and the divisor.
19. Li Shanlan, *Duoji bilei* (Analogical Categories of Discrete Accumulations), 4 *juan*, in *Zeguxizhai suanxue* (Mathematics from the Zeguxi Studio), Jinling ed., Vol. 4 (1867).
20. See Horng Wann-Sheng, *Li Shanlan: The Impact of Western Mathematics in China During the Late 19th Century*, PhD thesis, City University of New York (1991), p. 206. Renshu was the style name of Li Shanlan.
21. See [19, *juan* 2, p. 3a].
22. See [20, p. 225], 'one of the early masterpieces of combinatorics'. See also J.-C. Martzloff, Un exemple de mathématiques chinoises non triviales: Les formules sommatoires finies de Li Shanlan (1811–1882), *Revue d'Histoire des Sciences* 43(1) (1990), 81–98, and J.-C. Martzloff, *A History of Chinese Mathematics* (corrected second printing of the first English edition of 1977), Springer (2006), 341–51.



The arithmetical triangle of al-Karajī (c.1007).

CHAPTER 3

Islamic combinatorics

AHMED DJEBBAR

Combinatorics, considered in its general sense as the collection of manipulations and the study of configurations [1], appeared relatively early in such areas of medieval Arab intellectual activity as astrology, lexicography, music, chemistry, and even philosophy. Following the resurgence of traditional disciplines (such as geometry, the theory of numbers, and astronomy) and the development of new ones (such as algebra, Indian calculation, and trigonometry), new combinatorial preoccupations then arose, linked to the study of theoretical questions.

Introduction

Analysis of the documents that have come down to us indicates that the practice of combinatorics went through two distinct periods. In the first, prior to the 12th century, combinatorics was confined to listing and counting procedures, either arithmetical or mechanical, which led to the modelling of a range of different problems and thus to general propositions or formulas applicable to them. The second period, which began perhaps in the second half of the 12th century, was a return to certain combinatorial preoccupations, together with a revival in linguistic studies. So we observe the emergence of true propositions (announced and demonstrated), some calculating procedures (with a table or by arithmetical formulas), and then the application of new tools to solve problems arising from different areas.

In this chapter we present known aspects of these various combinatorial practices, basing our presentation on research from the past three decades. However, our conclusions and conjectures cannot be other than provisional, given the very incomplete nature of the extant sources.

Practical combinatorics before the 12th century

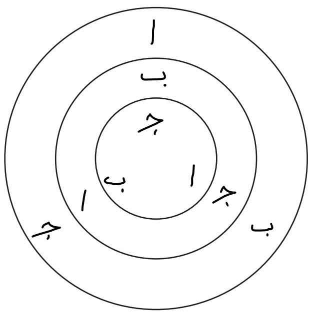
One of the most ancient disciplines that allowed for combinatorial manipulation was astrology. The need to know different configurations of planets, so as to use them in forecasting events, led naturally to their enumeration [2]. There were also astrologers who favoured the study of magic squares and circles [3] and constructed, or had constructed for them, numerous planar configurations that they called 'numbers in harmony': in the 10th century we find instances in the famous encyclopedia of Ikhwān aṣ-Ṣafā' [4]. But as far as the manipulation of integers was concerned (depending on arithmetic, rather than combinatorics), magic squares proved particularly interesting to mathematicians such as Abū l-Wafā' (d.997) [5], Ibn al-Haytham (d.1039) [6], and others after them.

With the coming of Islam, the privileged status of Arabic favoured the development of several 'sciences of language'. In this setting lexicographers listed (and sometimes counted) configurations of letters of the alphabet under certain constraints, with the aim of making glossaries. We know, for example, that al-Khalīl Ibn Aḥmad (d.786) gave the numbers of combinations of two, three, four, and five of the twenty-eight letters of the Arabic alphabet, and that the grammarian Sībawayh (d.796) subsequently determined the numbers of arrangements of these same letters, but taking into account incompatibilities of pronunciation (see [7]).

This combinatorial tradition was maintained by the linguists of the following centuries, but with some variations and in the setting of new preoccupations; we can cite, in particular, a book of Ḥamza al-Isfahānī (d.970), who repeated the counting effected by al-Khalīl (see [8]). After him, Ibn Jinnī (d.1005) included different arrangements of Arabic letters in his *Theory of Derivation*, trying to associate the meanings of all permutations of this combination with each three-letter combination having a particular meaning (see [9]). In Spain, in his *Mukhtaṣar* (Summary), az-Zubaydī (d.989) considered countings that take account of the constraints linked to pronunciation and usage.

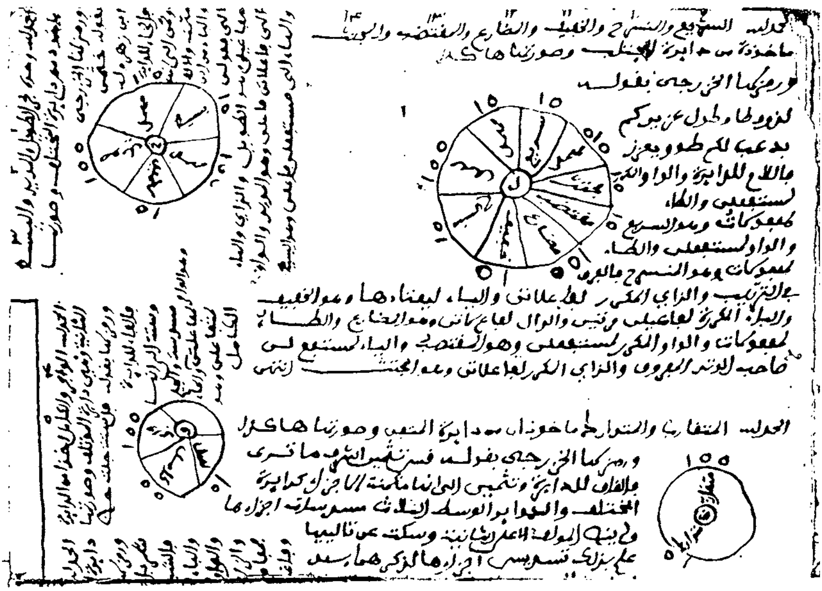
We note, moreover, on reading work by the polygraph as-Suyūṭī (d.1505), that even if the methods of calculation varied from author to author, they were all submitted to linguistic constraints (concerning rules of pronunciation or the nature of the letters forming the words) that might not lead to the formulation of general algorithms. In addition, these methods were not free of error, at the level of results and the reasonings that justified them. This all tends to prove that they did not yet have formulas or procedures to enable the counting.

This is also confirmed, for example, in the *Jamhara* of Ibn Durayd (d.933), where two different methods of calculation were incompletely presented. The first consisted of taking the letters to be combined, arranging them in two concentric rings, and then turning the rings to make different letters correspond. This procedure is identical to one that would later be found in astrology: Al-Būnī (d.1205), Ibn ᶜArabī (d.1240), and even Ramon Llull (d.1316), used it (see [10]).



Rotating rings of Ibn Durayd.

The second method consists of counting separately the words without repeated letters, and those with repetition, and counting the first group so as to distinguish between those words without *wāw* and *yā* and the rest. The calculation is correct for the arrangements with repetition of the twenty-eight letters taken in pairs, but is false for arrangements of more than two letters. The error shows up in the values of combinations of three letters: for example, Ibn Durayd seems to have used the formula $C(n,3) = n\,C(n,2)$ instead of $C(n,3) = \frac{1}{3}(n-2)\,C(n,2)$. It is astonishing that successive authors did not compare these results with those of al-Khalīl Ibn Aḥmad, who did not explicitly specify any particular method, but whose results were rigorously exact (see [11]).



Combinatorial circles of al-Khalīl.

In the area of chemistry Jābir Ibn Ḥayyān (8th century), one of the first great Arabic specialists in this field, theorized a kind of combination of elements constituting matter. Starting from the four elementary qualities that form the basis of Arabic chemistry and medicine (heat, cold, dryness, and humidity), he introduced the notion of degree, each comprising seven divisions (minutes, seconds, etc.). He then split the Arabic alphabet into four categories, corresponding to the four qualities, and established a direct relationship between the combinations of letters and those of the qualities. According to this logic, chemistry became a morphology of metals, in the same way that language is a composition of words (see [12]).

Music theorists also used elementary combinatorial procedures in their study of notes and scales. This was done particularly by al-Fārābī, who said, with regard to combinations of intervals [13]:

> When the intervals are all unequal, we can make three combinations. In the first, the largest of the three intervals is placed at one end, the smallest at the other. In the second, the largest is at one end, with the smallest at the centre. In the third, the largest interval is in the centre. And for each of these combinations, one can arrange the intervals either from the flat to the sharp, or from the sharp to the flat.

We find similar formulations regarding the combination of sounds in the works of Ikhwān aṣ-Ṣafā' (10th century) [14]:

If you arrange these three fundamental sounds in pairs, nine melodies will result . . . In triples there are ten combinations. Here we have all the types of melodies composed [from] the [fundamental] sounds: three of them [are] simple, nine are binary, and ten are ternary.

In the mathematical domain, the combinatorial approach that we observe is of the same nature as in other disciplines – that is, it did not require the formulation of general propositions to resolve problems. Combinatorial procedures frequently appeared in work that led to the resolution of problems that were harder and of a different nature. During the period that interests us here, as we shall see through various examples, it was astronomy and algebra that allowed mathematicians to confront problems whose solutions could not be obtained by classical arithmetical procedures.

Combinatorial approaches in astronomy

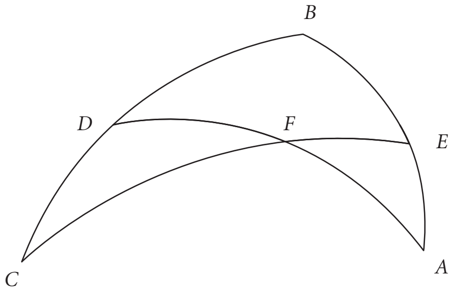
It was through trigonometrical problems that the elements of combinatorics (its approach and its results) arose in a chapter of Arabic astronomy which, in any case, concerned classifying problems according to their solubility and generalizing methods of calculation. From this dual perspective, precise calculations or simple enumerations enabled the classification of all cases considered as to their solvability and their methods of solution. Two works illustrate this well: the *Kitāb ash-shakl al-qaṭṭā*ᶜ (Book of the Secant Figure) of Thābit Ibn Qurra (d.901) and the *Kitāb Maqālīd* ᶜ*ilm al-hay'a* (Book of the Keys to Astronomy) of al-Bīrūnī (d.1048).

Beyond his trigonometrical preoccupations, Ibn Qurra's treatise is essentially dedicated to the solution of the following combinatorial problem:

To count, list, and justify all the possible ways of writing, the problem of the 'secant figure'

– that is, the formula given by Menelaus in his *Sphaerica* (Spherics), Book III, Proposition 1:

$$\frac{\text{chord}(2AE)}{\text{chord}(2EB)} = \frac{\text{chord}(2AF)}{\text{chord}(2FD)} \times \frac{\text{chord}(2DC)}{\text{chord}(2CB)}.$$



If we denote the six elements of this equality by $a_1, a_2, \ldots, a_6$, we have

$$\frac{a_1}{a_2} = \frac{a_3}{a_4} \times \frac{a_5}{a_6} \text{ or, equivalently, } a_1 a_4 a_6 = a_2 a_3 a_5.$$

From this formula, Ibn Qurra showed that one could deduce several other relationships, which he listed and whose validity he demonstrated. First he obtained eighteen permutations, from which he proved the existence of a relationship analogous to the above formula and which he enumerated in a table [15]; then he deduced eighteen other permutations from the first by permuting the numerators and denominators of each fraction. He concluded by showing that only the thirty-six permutations obtained correspond to the problem. For this, he considered the fifteen combinations of six elements, $a_1, a_2, \ldots, a_6$, taken in pairs [16]. He studied nine of these, leaving out the six following combinations:

$$(a_1; a_4),\ (a_1; a_6),\ (a_2; a_3),\ (a_2; a_5),\ (a_3; a_5),\ (a_4; a_6).$$

As he showed, these combinations have to be excluded because they form relationships that do not arise from the secant figure [17].

It was also a problem of spherical trigonometry that allowed al-Bīrūnī to have recourse to combinatorial methods. Like Ibn Qurra, he solved specific problems without reference to rules or to previous results, and like him, he explicitly used the general concept of combinations with (moreover) an identical formulation, in spite of the different nature of the combinations dealt with by the two writers.

This is how he proceeded. After expounding his own proof of the relation arising from the theorem of sines, he proposed to establish a classification of spherical triangles, and then to indicate how to calculate the unknown elements from the known ones,

the goal being to determine, where possible, all the elements (sides and angles) of a spherical triangle when given just one, two, or three of them (see [18]). In this, he counted the triangles according to the nature of the interior angles formed by the three sides to determine the combinations with repetition of three angles, taken three at a time.

Much later, in his *Kitāb ash-shakl al-qaṭṭā*ᶜ (Book of the Secant Figure), Naṣīr ad-Dīn aṭ-Ṭūsī (d.1274) revealed a more systematic and detailed study of spherical triangles, with supplementary combinatorial considerations – for example, combinations with repetition of three types of sides of a triangle, and a figure of compatible combinations of ten types of angles (already studied by al-Bīrūnī) and ten types of sides (see [19]).

Combinatorial approaches in algebra

It was through systems of equations with integer solutions that combinatorial approaches made their appearance in algebra. In one of his extant works, Abū Kāmil (d.930) wished to show the existence of problems whose solutions were very large numbers. Each time, he was led to make an exact count of the solutions of the system being studied. But since the search for integer solutions of each of these systems is equivalent to a counting problem with constraint, only the enumeration of the solutions was accessible to the author. In effect, he was led each time to counting the quadruples (x, y, z, t) in a set A defined by

$$A = \{(x, y, z, t): \ y = ak, z = bm, t = cn, x = g(y, z, t), x + y + z + t < d\},$$

where a, b, c are fractions, k, m, n are positive integers, and g is a linear relation. The inequality that appears in each problem corresponds to the constraint that complicates the counting, and that moved Abū Kāmil away from the lexicographical procedures familiar to linguists. An example is given below.

After Abū Kāmil, combinatorial approaches emerged, both implicitly and explicitly, in the contributions of al-Karajī (d.1029) and as-Samaw'al (d.1175). The former constructed the arithmetical triangle inductively, so as to use its coefficients (in any order) for the development of the binomial, in order to study some mathematical operations applied to polynomials. But, probably because of the nature of this purely algebraic approach, al-Karajī did not separate out the

Abū Kāmil's problem on the buying of birds [20]

This problem concerns buying, with 100 dirhams (100d), 100 birds of four types: geese at 4d each, chickens at 1d each, pigeons at 1d for two, and starlings at 1d for ten:

$$x = \tfrac{3}{10}y + \tfrac{1}{6}z.$$

From this,

$$A = \{(x, y, z, t)\colon\ x = \tfrac{3}{10}y + \tfrac{1}{6}z \quad \text{and } t = 100 - x - y - z\}.$$

This splits A into two subsets, with

$$B = \{(x, y, z, t) \in A\colon\ y = 10n, z = 6k\}$$

and

$$C = \{(x, y, z, t) \in A\colon\ y = 10(n + \tfrac{1}{2}); z = 6(k + \tfrac{1}{2})\}.$$

But $B = \bigcup B_k$, with

$$\begin{aligned} B_k &= \{n \geq 1\colon\ (3n + k) + 10n + 6k < 100\} \\ &= \{n \geq 1\colon\ n < \tfrac{1}{13}(100 - 7k)\}, \end{aligned}$$

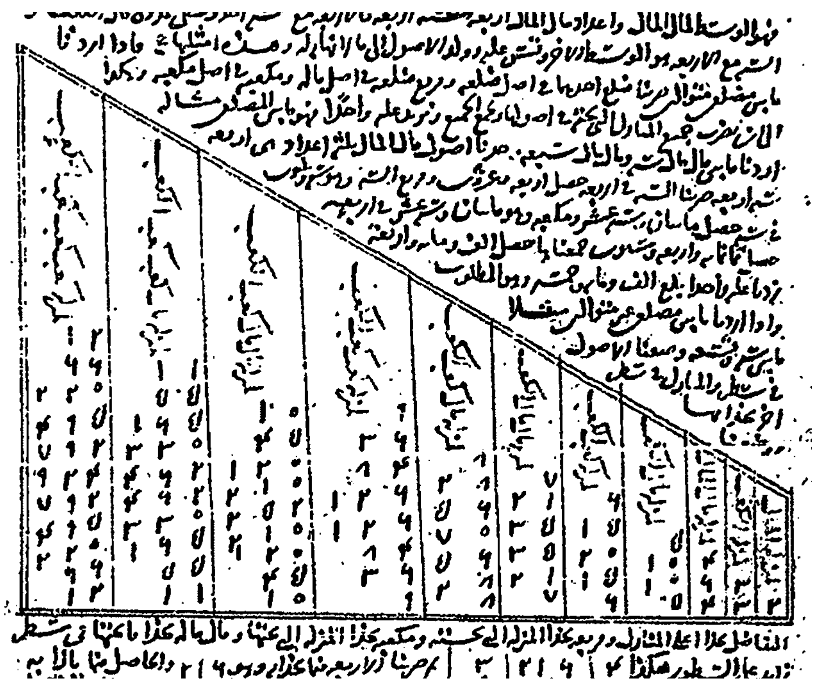
and $C = \bigcup C_k$, with

$$\begin{aligned} C_k &= \{n \geq 0\colon\ 3(n + \tfrac{1}{2}) + (k + \tfrac{1}{2}) + 10(n + \tfrac{1}{2}) + 6(k + \tfrac{1}{2}) < 100\} \\ &= \{n \geq 0\colon\ n < \tfrac{1}{13}(90 - 7k)\}. \end{aligned}$$

To determine $|B_k|$ for $1 \leq k \leq 12$, and $|C_k|$ for $0 \leq k \leq 11$, the author proceeded by enumeration;
finally, he obtained

$$|A| = |B| + |C| = \big(|B_1| + \cdots + |B_{12}|\big) + \big(|C_1| + \cdots + |C_{11}|\big) = 45 + 53 = 98.$$

combinatorial significance of the coefficients which he calculated with the aid of his figure. The second aspect of the arithmetical triangle is similarly absent from the work of as-Samaw'al, in which the latter presented certain contributions of his predecessor and developed them with original results. It is totally possible that al-Bīrūnī [21] may have independently used this same arithmetical triangle in one of his studies on the extraction of the *n*th root of a number, but we have no confirmation of this because his treatise has not come down to us. We know that al-Khayyām (d.1131) wrote a book, still undiscovered, in which he presented a procedure for the extraction of the *n*th root of a number, based, it would seem, on the development of the binomial (see [22]). The same triangle is found much later in a work on calculation by Naṣīr ad-Dīn aṭ-Ṭūsī [23].



The arithmetical triangle of Naṣīr ad-Dīn aṭ-Ṭūsī.

Also in relation to the development of algebra, a combinatorial approach featured in the *Kitāb al-bāhir* (Flamboyant Book) of as-Samaw'al. In one chapter, the author proposed to classify problems in terms of their solubility, the number of their solutions, and conditions for compatibility. He began by distinguishing between the problems that are possible and those that are impossible. But he also made a second classification, according to the finite or infinite number of solutions of the given equation. To illustrate a third classification, in terms of conditions for compatibility, he enumerated a set of equations. As an example, he proposed to find ten integers, six of which add up to a given number (see [24]). To determine the $C(10, 6) = 210$ solutions, he proceeded to a systematic enumeration of all the equations. Then he obtained, always by enumeration, the number of conditions for compatibility, which is $9 \times C(8, 5)$.

As we have stated, the problems of a combinatorial nature solved in the examples that we have just presented all have a common feature: the absence of an explicit reference to combinatorial results outside astronomy and algebra that can serve as models. We also note that, contrary to astronomy, no element of combinatorial terminology features in the manipulations of the algebraists whom we have just briefly presented.

Additionally, we must emphasize that we do not have enough information to be able to affirm that the combinatorial nature of the problems that appeared in the areas of astronomy and algebra, which we have just described, was noticed by mathematicians from the East and the West after as-Samaw'al. In any case, it appears very improbable that countings associated with the indeterminate analysis of Abū Kāmil, and the binomial coefficients of al-Karajī whose strictly algebraic solution completely hides the combinatorial aspect, were noticed elsewhere, particularly because of the absence of symbols.

Combinatorics in the East

We have very few sources for combinatorial practices in the Muslim East after the 12th century. But the documents that have come down to us show that these practices always involved the solution of particular problems in one or another discipline, and also show where the combinatorial approach was used as a tool for the solutions, without reference to previous results. Having said that, in view of the methods used and the results obtained, we can say that one sees a slow phenomenon of accumulation of practices that go objectively in the direction of the emergence of a new chapter in mathematics.

In philosophy, one notices the introduction of combinatorial approaches in connection with the study of logical or metaphysical problems. This was the case, for example, with one of the questions treated by Ibn Sīnā (d.1037) in his *Kitāb ash-shifā'* (Book of Healing) and in his *Kitāb al-ishārāt wa* *t-tanbīhāt* (Book of Directives and Remarks). In the 13th century, Naṣīr ad-Dīn aṭ-Ṭusī took up the thoughts of Ibn Sīnā again, and proposed to elucidate a mathematical aspect. For that, he calculated the number of combinations of n elements, one at a time, two at a time, ..., n at a time, and deduced the sums and products of numbers of this type. Here, also, the author did not refer to combinatorial propositions that were established before him, particularly those to be found in Maghreb (north-west African) works of the 12th and 13th centuries, which we discuss at length in the following section (see [25]). This is also an example of the non-circulation of certain scientific writings, even within the Muslim Empire.

In mathematics, two oriental texts from after the 12th century survive, each of which, in its own way, bears witness to the existence and persistence of explicit combinatorial preoccupations and approaches. The older one is the *Tadhkirat al-aḥbāb* of al-Fārisī (d.1320), a scientist known particularly for his commentary on Ibn al-Haytham's *Book of Optics*. In his study of the decomposition of an integer as a product of prime factors [26], he was led to determine all the possible combinations of these numbers to obtain the divisors of the given integer. For this, he constructed a truncated arithmetical triangle and identified with each $C(n, k)$ the entry of the triangle to be found at the intersection of the kth row and the $(n-k+1)$th column. But at no point did al-Fārisī refer to an arithmetical triangle, which, as we have seen, had been constructed and used by a number of mathematicians between the 11th and 13th centuries. To help the reader represent the divisors of a number concretely from its decomposition into prime factors, the author also constructed a second table that contained the results of his enumerations: a first column for all the forms of decomposition of an integer into prime factors, a second column for the number of divisors arising from this decomposition, and a third column for examples corresponding to each decomposition.

The second mathematical text containing these countings is from the 15th century. It is *Ḥāwī l-lubāb* (The Collection of the Marrow) by the Egyptian Ibn al-Majdī (d.1447). It is a commentary on a celebrated Maghreb 14th-century manual, the *Talkhīṣ* (Summary) of Ibn al-Bannā (d.1321). In this treatise [27], the author studied a new problem that reveals incontestably a realization of, and interest in, combinatorial tools for the resolution of mathematical questions. In fact, he presented a method that allows the counting of all equations with positive coefficients that one can write with monomials of degree less than or equal to a given integer n; its approach rests on combinatorial reasoning and induction. But, as we shall see, the calculations and the countings approach of Ibn al-Majdī, which are not found in any other known work, seem to have been written in a more ancient combinatorial tradition whose main focus was the city of Marrakech in the far western Maghreb.

Combinatorics in the Maghreb: Ibn Munᶜim

According to the information that is available to us today, it was in northwest Africa that practical combinatorics started a new chapter, with its objects, tools, and areas of application. But, as in the East, it was a non-mathematical discipline, linguistics, that prepared the origins of this chapter. This activity was certainly not new to the Maghreb, but during the centuries that concern us it was about to benefit from a real dynamism (see [28]). In fact, from the 12th century we note a regaining of interest in the different chapters of linguistics that non-mathematicians dedicated themselves to studying. Moreover, it is not by chance that it was Ibn al-Bannā, the author of several works on the Arabic language [29], who was one of those writers whose combinatorial preoccupations had the greatest consequences.

To be precise, we note that problems were posed and solved by using formulas and reasoning of a combinatorial character, that this terminology, born from the needs of linguists, acquired a mathematical status, and that a new form was established to be used as an instrument operating on mathematical objects. But without a detailed knowledge of the different aspects of this activity since its origins, we limit ourselves here to pointing out and enlarging upon certain salient features.

From the end of the 12th century or the beginning of the 13th, some problems posed by Arabic linguistics of the 8th century were re-examined, this time within the framework of an autonomous chapter on the science of calculation. This is what Ibn Munᶜim, a mathematician from Marrakech, did in his treatise *Fiqh al-ḥisāb* (Science of Calculation). But before we present the details of his contribution, it is useful to say a few words about this relatively little-known mathematician, who died in 1228.

Modern bibliographers, such as H. P.-J. Renaud, H. Suter [30], and others, have confused this mathematician with the geometer and astronomer Muḥammad Ibn ᶜAbd al-Munᶜim, who lived in Sicily at the court of Roger II.

However, their common source was the historian Ibn Khaldūn, who did not mention Sicily when he spoke of this mathematician [31]. Amongst the ancient Maghreb biographies that we could consult, only one mentions him – Ibn ᶜAbd al-Malik (d.1303), who gives us the following information. Ibn Munᶜim came from Denia, on the coast of Spain near Valencia. He stayed in Marrakech, where he taught. He was known at the time as one of the best specialists in geometry and number theory. At the age of 30, he started studying medicine, which he practised later with success while still pursuing his mathematical activities. As well as his work *Fiqh al-ḥisāb*, he published writings on geometry and magic squares, but none of his works on these two disciplines has come down to us.

It is in the 11th section of his book, entitled 'the counting of words that are such that the human being can only express himself by one of them', that Ibn Munᶜim described his approach and his combinatorial results. The contents of this section were presented by the author as an extension of the results of al-Khalīl Ibn Aḥmad, the great 8th-century linguist, and is a generalization of his calculations which allowed him to determine combinations without repetition of Arabic phonemes (see [32]). To this effect, Ibn Munᶜim proposed to treat the problem first in a general manner, and then to follow it by demonstrations of examples and pictures. However, as the analysis of this section confirms, the generality of which the author spoke did not leave the fixed linguistic setting, and was concerned with the establishment of mathematical formulas and procedures with a view to counting words of any length in any language. Despite all this, his study objectively passed beyond the linguistic setting in which it was formulated and realized, as much in the manner of posing the problems and linking them together by the methods of reasoning used as by the results established.

Starting with a set of silk colours that played the role of the abstract model, Ibn Munᶜim established a rule allowing one to determine all the possible combinations of n colours, k at a time. For this, he used an inductive method to construct a numerical table, where he identified the elements with the combinations he sought. Using a strictly combinatorial approach he thereby gave, for the first time to our knowledge, the famous arithmetical triangle that the algebraists at the centre of the (Islamic) Empire had already constructed in an algebraic setting, with a view to determining binomial coefficients.

The arithmetical triangle of Ibn Munᶜim.

It is to be regretted that, fearing 'excesses and length', Ibn Munᶜim decided against expounding the 'extraordinary properties' (as he put it) that he knew how to obtain from a simple comparison of elements in the table. Given the ability of this mathematician, it is certainly possible that he had noticed, on a single reading of the table of numbers, new relationships between its positions, and thus new combinatorial formulas. It is equally possible that he contented himself with just showing it to his students. From later sources, we know that one of his students pursued his teachings and inspired in his turn some new mathematicians. This was al-Qāḍī ash-Sharīf (d.1283), whose book *al-Qānūn fī l-hiṣāb* (Canon of Calculation) has not yet been found. Its contents could throw light on the eventual place of combinatorics in mathematical teaching in Marrakech in the 13th century.

The studies of Ibn Mun^{c}im were pursued, following a combinatorial approach based on induction, by establishing formulas relating to permutations (with or without repetition) of a set of letters, and formulas given recursively, for the number of possible readings of the same word of *n* letters (taking into account all the pronunciation signs used in a given language).

The author concluded this first part by establishing the formula for arrangements without repetition, adopting a process analogous to the preceding one and necessitating having recourse to tables of numbers (see [33]). As in his research into simple combinations, Ibn Mun^{c}im once again made his sets of objects (such as letters of the alphabet, colours of silk, or coloured threads) function like abstract models by identifying each time the object being studied with the elements of the model. Taking into account the approach he adopted, Ibn Mun^{c}im was also led to tackle problems on partitions of numbers. However, at no time did he mention this question in its full generality.

As well as several applications, the third part of the 11th section contains a series of tables that permit us to determine one at a time all the elements that appear in the counting of words that one can pronounce in a given language – that is, the combinations, permutations, and arrangements, with or without repetition.

So as to emphasize the 'theoretical' aspect of his study, Ibn Mun^{c}im stated firmly that the tables in the last part of the section did not exist in the first version of the book. He added these

> after the work was recopied and it was in the hands of students

with the aim of illustrating the general procedures established in the first two parts.

It remains for us to say a few words about the types of reasoning used by the author to establish his results, and about their status with regard to tools and traditional mathematical approaches. An analysis of the author's approach

Results established by Ibn Munʿim

(1) Combinations in the arithmetical triangle:

$$C(n,k) = C(k-1,k-1) + \cdots + C(n-1,k-1).$$

(2) The number of permutations in a word of n distinct letters:

$$P_n = 1 \times 2 \times 3 \times \cdots \times n.$$

(3) The number of permutations in a word of n letters of which p letters are repeated (respectively) $k_1, k_2, \ldots, k_p$ times:

$$P(n,k) = p_n/(p_{k_1} \times p_{k_2} \times \cdots \times p_{k_p}).$$

(4) The number of pronunciations of a word, taking note of vowels:

$$S_n = 4S_{n-1} - 3S_{n-3}, \text{ or equivalently } S_n = 3S_{n-1} + 3S_{n-2}.$$

makes two types of reasoning apparent, which one could qualify globally as *inductive* and *combinatorial*. The first of these, with its different variants (and with the meaning that it had up to the 17th century), is a traditional tool in Arabic mathematics, with its privileged area (the theory of numbers) having a particular, but recognized, status. For the second type of reasoning, which we have not seen in mathematical writings before the *Fiqh al-ḥisāb*, induction was distinguished by an approach that combined enumeration and counting. The use of this approach to establish rules that we consider general seems to be a factual recognition of its mathematical character. However, we cannot yet affirm that an explicit recognition had truly taken place, in spite of the presence (after the 13th century and in many other teaching texts) of reasonings and of combinatorial propositions already established and operating like new tools.

The presence of all these elements (approaches, results, commentaries) in the same mathematical work certainly begs the question of the degree of originality of the author. In fact, it seems difficult to attribute to one single mathematician at the same time the announcement of a number of new results (by comparison with the known Arabic tradition), the mathematization of these results, as much in their formulation as in their establishment, and above all, their introduction as an autonomous subject in a work dedicated essentially to the science of calculation and the theory of numbers. At the beginning of his book, Ibn Munᶜim stated specifically that he had not been content to report the results and proofs of the Ancients, but their unfortunately stereotyped and over-vague formulation added to our ignorance of certain aspects of the history of combinatorics between the 8th and 9th centuries. Consequently, we are unable to appreciate the nature and importance of the personal contribution of Ibn Munᶜim.

That being so, when compared (for example) with the fumblings and the absence of rigour that one finds in certain combinatorial problems up to the 17th century, the elaborate character of the results and approaches present in the *Fiqh al-ḥisāb* and the spirit of method that was released force us to move the beginning of the mathematization of this discipline within the framework of Arabic science to well beyond the epoch of Ibn Munᶜim. In the absence of new information to prove or disprove this hypothesis, the *Fiqh al-ḥisāb* remains the most ancient known work in which the double aspect of theory and practice is included in a chapter devoted to combinatorial analysis.

But his importance does not stop here. First of all, and with regard to the Arabic linguistic tradition, his work seems like a culmination in the extent to which he presented a general solution to the problem posed by the first Eastern lexicographers. Secondly, on a strictly mathematical plane, this work represents an important stage reached where, as we can see in detail, it marks the end of a series of combinatorial practices, that of calculation by enumeration or by means of tables, and the beginning of another stage, that of the extension of 'formulas' and the enlargement of their sphere of use.

Combinatorics in the Maghreb: Ibn al-Bannā

Towards the end of the 13th century at the latest, a new threshold was crossed in combinatorial activity in the Maghreb. Formulas expressing the number of combinations and arrangements of n objects taken k at a time were given, but with new proofs and in the setting of a classical area, that of the theory of numbers. This contribution was presented and elucidated explicitly by Ibn al-Bannā in two of his works, the *Tanbīh al-albāb* (Warning to Intelligent [People]) and the *Rafᶜ al-ḥijāb* (Lifting of the Veil).

The first of these was a collection of problems, coming mainly from areas outside mathematics. In Problem 14, entitled 'A question taken from linguistics', he first announced the arithmetical rule, already given by Ibn Munᶜim, that permits the calculation of permutations of any number of letters of the alphabet. But what is new is what he said about the counting of combinations. In fact, not only did he give (for the first time to our knowledge) the arithmetical formula that allows explicit calculations of combinations without repetition ($C(n, k)$, for $n \geq 2$ and $2 \leq k \leq n$), but he also took care to mark himself out from his predecessor Ibn Munᶜim and the latter's method based on the construction of the arithmetical triangle. Here is what he said on the subject [34]:

> As far as [knowing] how many 3-letter or 4-letter words can be composed from the twenty-eight letters of the dictionary, Ibn Munᶜim made a table for this. And I have never seen anybody simplify [the problem] by a rule. I have therefore thought about this and the idea of a simple procedure came to me: this consists in considering [successive] numbers which differ by 1 and whose number is equal to the number of letters of the combination and whose greatest is 28. Then, you consider the successive numbers, starting from 1, whose number is equal to the number of letters of the combination. Then you divide what results from the product of the first numbers, the ones by the others, by what results from the product of the other numbers. What results from this is the number of words. And one must suppress the [numbers] common to the [number] divided and to the divisor before the multiplication of their numbers, one at a time, so that it becomes easier and shorter [to calculate].

Thus we obtain the classic formula

$$C(n, k) = \frac{n \times (n-1) \times \cdots \times (n-k+1)}{k \times (k-1) \times \cdots \times 2 \times 1}.$$

The *Rafᶜ al-ḥijāb* was conceived by Ibn al-Bannā as a partial commentary on, and complement to, his manual *Talkhīṣ aᶜmāl al-ḥisāb* (Summary of Methods) for the *Tanbīh al-albāb* [35]. It is therefore not surprising that we rediscover there the results that we have just discussed. But this time they are presented in a more general classical setting, that of number theory in the Pythagorean tradition such as it travelled to the Maghreb via the Arabic translation of the *Introduction to Arithmetic* of Nicomachus (2nd century). In fact, the combinatorial expressions are compared to the sums of finite sequences of integers and of the elements of the table of figurate numbers.

This explicit arithmetic approach enables us to say that Ibn al-Bannā had a clear perception of the close link, at the level of both results and proofs, between the methods of arithmetic and combinatorics. For results, we can state specifically that it was the table of polygonal numbers that affirmed the link between the two disciplines. But for proofs, there are different methods of induction that justify Ibn al-Bannā's integration of combinatorial results into the vast topic of number theory. In a more precise way, Ibn al-Bannā used the procedure of *regression* to establish the sums of the finite series of n terms and an induction of 'almost general' type, operating this time on the propositions of double index $P(i, j)$ for any j (see [36]).

The application of combinatorial results

As well as the formulas attributed to him, Ibn al-Bannā seems to have been the first in the Arab tradition to become interested in the combinatorial aspects of certain classical problems. In this way, in the *Rafᶜ al-ḥijāb*, he took up Thābit Ibn Qurra's work on the secant figure by using combinatorial reasoning to simplify it considerably: all the forms of Menelaus's theorem were obtained as results of arrangements arising from the expression $a_1 a_4 a_6 = a_2 a_3 a_5$ (see [37]).

In the same treatise Ibn al-Bannā counted the different equations that result from a geometric or arithmetic progression, when we consider each of these as a set of elements characterizing it. An arithmetic progression was thus identified with a set of five elements: the first term u_1, the last term u_n, the number of terms n, the common difference r, and the sum S_n. The different partitions of this set into two subsets containing the 'knowns' and the 'unknowns' correspond to the number of equations sought. To determine this number, he used (without saying so) the formula for combinations of n objects taken k at a time [38]. This is exactly the same process that he adopted in his little book on elementary geometry entitled *Risāla fī t-taksīr* (Epistle on Measuring) [39], where the classic geometrical figures are identified with the set of their constituent elements (angles and sides). Finally, in his book known as *Arbaᶜ maqālāt* (The Four Epistles) [40], he counted the different combinations of integers and fractions linked by the elementary arithmetical operations.

These countings of very different objects were possible through his clear perception of the correspondence established between a finite set of any kind and a subset of the alphabet playing the role of the abstract set. It was therefore one step further in mathematical symbolization because, from then on, to operate on any objects it was sufficient to manipulate the letters. Ibn al-Bannā, and later, Ibn Haydūr (d.1413) in his *Jāmiᶜ*, said so explicitly [41].

It was also Ibn al-Bannā who was the first to have the idea of reassembling and solving non-mathematical problems with a combinatorial aspect. In his *Tanbīh al-albāb*, he dealt with the following problems:

- enumerating the different cases of possible inheritance when the inheritors are n boys and k girls (Problem 1);
- announcing all situations where washing with water is necessary, and those where one can dispense with water (Problem 2);
- enumerating, according to the strictures of the Malekite rite, all prayers for compensating for forgetting some of them (Problem 4);
- enumerating all possible readings of the same phrase, according to the rules of Arabic grammar (Problem 15) [42].

After Ibn al-Bannā, certain commentators who had assimilated the earlier contributions extended them or applied them to new problems. This was the case with Ibn Haydūr in the Maghreb and Ibn al-Majdī in Egypt [43]. We note that the presence of the same vocabulary used by these different authors, and the fact that none of them explicitly revised the known results, reinforced the continuity of combinatorial preoccupations, at least since Ibn Munᶜim.

Taking into account the quantitative and qualitative importance of the results and practices that we have just described, we naturally ask ourselves about the presence of combinatorial elements in the mathematical teachings of the Maghreb from the 13th century onwards. The appearance in the works of several commentators of combinatorial approaches and a preoccupation with the *Talkhīṣ* of Ibn al-Bannā is a prime argument in favour of this hypothesis. Ibn Khaldūn gave us a second argument: concerning mathematical commentaries in the paragraph of his *Muqaddima* on linguistics, he reproduced the classic results (transmitted since al-Khalīl) on combinations and arrangements of the letters of the alphabet. But he also added two proofs, perhaps taught to him by his teacher al-Ābilī, who himself was a pupil of Ibn al-Bannā. Moreover, in one of them, which allows the establishment of the combinations of *n* letters of the alphabet taken two at a time and three at a time, he committed an error of reasoning (and not of calculation) that could betray a difficulty in assimilating combinatorial approaches (see [44]).

Conclusion

The different elements presented in this chapter show, we believe, that new material and new tools seem to have emerged slowly in the setting of the Arab scientific tradition, by the solution of concrete problems and then their continued mathematization. That said, we do not know how much understanding these people had, nor how much importance they gave to these developments. In any case, this did not go so far as to give a name to this activity and to distinguish it from classical operations on integers, despite the use of these results in other areas of mathematics. Combinatorics did not thus benefit from the favourable conditions of the beginning of the 11th century that enabled the first algebraic practices to be transformed into a discipline with a name, a status, and a wide range of applications. Moreover, combinatorics did not have the good fortune of trigonometry, which was descended from astronomy but which succeeded in detaching itself towards the middle of the 11th century with the publication of works (those of al-Bīrūnī and of Ibn Muᶜādh) dedicated exclusively to it (see [45]).

Among the causes that could explain the limited development of combinatorics in Islamic countries, there was first the absence of local or regional institutions charged with renewing programmes and imposing, and then perpetuating, the teaching of new ideas. Secondly, there was the influence of the general environment: as we have already said, combinatorics seems to have been taught until the 13th century in Marrakech, but the decline of scientific activity in the 14th century (witness Ibn Khaldūn in his *Muqaddima*) did not spare the theoretical areas of mathematics. All of this became concretized in a cessation of research and a decreased interest in anything that did not immediately concern the applications of science.

There is finally the nature and status of the areas of application of combinatorics. In the 14th century, linguistics no longer provided a source of problems and inspiration for practical use and reflections of a combinatorial nature, because the problems had been completely solved by mathematicians. Philosophy also no longer had the vitality it had in the 9th to the 12th centuries, and offered no new problems. There remained just one area, that of games of chance, which despite religious prohibition could be a substitute and source of inspiration for practical combinatorics, but so far no Arab texts on this theme have been discovered.

In fact, combinatorial practices needed to continue to develop theoretically in a different scientific and cultural context, thereby opening the field to new types of problems. This would be done in the Europe of the 17th century.

Notes and references

1. The definition that we adopt here is explained and justified in C. Berge, *Principles of Combinatorics*, Academic Press (1971), 1–10. It is both more general and more fruitful than those given in G. Papy's *Modern Mathematics*, Vol. V, p. VI, or N. Bourbaki's *Elements of Mathematics*, Vol. XX, Book I, Chapter 3, pp. 62–6.
2. As a classical example of combinatorics, the 14th-century Maghreb mathematician Ibn Haydūr gave the conjunctions of seven planets, two at a time, three at a time, . . . , six at a time, in his *Tuḥfat aṭ-ṭullāb wa umniyyat al-ḥussāb fī sharḥ mā ashkala min Rafᶜ al-ḥijāb* (The Finery of Students and the Desire of Calculators in the Explanation of Difficulties in the Lifting of the Veil), Vatican, Ms. 1403, f. 52b.
3. The classical magic squares of order n are plane configurations consisting of the first n^2 whole numbers, arranged in n rows and n columns in such a way that the sum of the elements in each row, each column, and each main diagonal is constant and equal to $\frac{1}{2}n(n^2+1)$ (see Chapter 1). For magic circles, see al-Būnī, *Shams al-maᶜārif al-kubrā* (The Sun of More Great Learning), Beirut (undated). For a detailed knowledge of the Arabic tradition of magic squares, see J. Sesiano, An Arabic treatise on the construction of bordered magic squares, *Historia Scientiarum* 42 (1991), 13–31, and Quelques méthodes arabes de construction des carrés magiques impairs, *Bull. Soc. Vaudoise des Sciences Nat.*, 83 (1994), 51–76.
4. Ikhwān aṣ-Ṣafā', *Rasā'il* (Epistles), Vol. I, Beirut (1957), 109–13.
5. Abu l-Wafā', *Risāla fī tarkīb ᶜadad al-wafq fī l-murabbaᶜāt* (Epistle on the Construction of Harmonic Numbers in Squares), Ayasofya Ms. no. 4843/3°, ff. 23b–56b.
6. Ibn al-Haytham, *Maqāla fī aᶜdād al-wafq* (Memoir on the Harmonic Numbers), cited by F. Sezgin, *Geschichte des Arabischen Schrifttums*, Vol. V, Brill, Leiden (1974), 372.
7. Ibn Khaldūn, *al-Muqaddima* (Introduction), Vol. I, Dār al-Kitāb al-Lubnānī, Beirut (1967), 1250–6.

8. F. Rosenthal, *An Introduction to History*, Vol. III, p. 327 (1958), note 1260.
9. A. Mehiri, *Les Théories Grammaticales d'Ibn Jinnī*, Tunis (1973), 250–67.
10. A. Djebbar, *Enseignement et Recherches Mathématiques dans le Maghreb des XIIIe–XIVe Siècles*, Université Paris-Sud, Publications Mathématiques d'Orsay (1980), no. 81–01, p. 135, note 222.
11. as-Suyūṭī, *al-Muzhir fī ᶜūlum al-lugha* ([Book] Flourishing on the Sciences of Language), Cairo, undated edition, pp. 71–6.
12. P. Kraus, *Jābir Ibn Ḥayyān, Contribution à l'Histoire des Idées Scientifiques dans l'Islam*, Cairo (1942–3).
13. al-Fārābī, *Kitāb al-mūsīqī al-kabīr* (transl. R. D'Erlanger), Paris (1930), 59–60.
14. Ikhwān aṣ-Ṣafā' (see Note 4, p. 198).
15. Ibn Qurra, *Kitāb fī ash-shakl al-mulaqqab bi l-qaṭṭāᶜ* (Book on the Secant Figure), Alger. B. N., Ms. no. 1446, ff. 83b–94a. This problem was taken up again by the same author in a work on connections, *Risāla ilā al-mutaᶜallimīn fī n-nisba al-mu'allafa* (Epistle to Those Students Who Study Connections), Paris Ms. no. 2457/15, ff. 60b–75b.
16. Here are the terms that Ibn Qurra used: 'If we combine one of six sizes, and if each remaining size is connected to one already counted, the result is fifteen combinations' (see Note 15, *Kitāb*, f. 92b).
17. Ibn Qurra says on this subject: 'The set of figures is numbered 18 and their permutations, that is all – no more and no less for the six combinations, as we have shown earlier' (see Note 15, *Kitāb*, f. 94a).
18. M.-T. Debarnot, *Kitāb maqālīd ᶜilm al-hay'a* (Book of the Keys of Astronomy), *La Trigonométrie Sphérique chez les Arabes de l'Est à la Fin du X^{e} Siècle*, Publications of the Institut Français of Damascus (1985).
19. A. P. Carathéodory, *Traité du Quadrilatère*, Constantinople (1891), 93–105.
20. Abu Kāmil, *Kitāb aṭ-ṭarā'if fī l-ḥisāb* (Book of Rare Things in Calculation), Paris B. N., Ms. no. 4946, ff. 7b–l0a.
21. al-Bīrūnī, *Kitāb fī-istikhrāj al-kiᶜāb wa aḍlāᶜ mā warā'ahū min marātib al-ḥisāb* (Book on the Extraction of the Cubic [Root] and of the Roots Which Are Beyond It in the Order of Calculation).
22. A. Djebbar and R. Rashed, *L'Oeuvre Algébrique d'al-Khayyām*, I.H.A.S., Aleppo (1981), 20.
23. Naṣīr ad-Dīn aṭ-Ṭūsī, *Kitāb jawāmiᶜ al-ḥisāb bi t-takht wa-t-turāb* (Collection of Calculations with the Aid of Board and Dust), Escorial Library, Ms. no. 973.
24. S. Ahmed and R. Rashed, *al-Bāhir fī l-jabr d'as-Samaw'al* (Flamboyant Book in Algebra of as-Samaw'al), University of Damascus (1972), 232–46 (Arabic text).
25. A. Nurani, *Talkhīṣ al-muḥaṣṣal*, Tehran (1980), 509–15.
26. al-Fārisī, *Tadhkirat al-aḥbāb fī bayān at-taḥāb* (Memoir of Friends to Show an Amiable Relation), Istanbul, Ms. Köprülü 941/2e, ff. 131b–139a.
27. Ibn al-Majdī, *Ḥāwī l-lubāb* (The Collection of the Marrow), British Museum, London, Ms. Add 7469, ff. 193b–194a.

28. Ibn Khaldūn is situated in this Renaissance of linguistic studies, in Spain and in the 14th century: see *al-Muqaddima* (Note 7, p. 1288).
29. Among the writings of Ibn al-Bannā in the Arabic language, one can cite *al-Kulliyyāt fī l-ᶜarabiyya* (Collection on the Arabic [Language]) and *Risāla fī ṭabīᶜat al-ḥurūf wa munāsabatihā li l-maᶜnā* (Epistle on the Nature of Letters and Their Correspondence with Meaning). See also H. P.-J. Renaud, Notes critiques d'histoire des sciences chez les musulmans, II, *Hesperis* 25 (1938), 39–49.
30. H. Suter, *Die Mathematiker und Astronomen der Araben und ihre Werke*, Teubner, Leipzig (1900), 217: see H. P.-J. Renaud, Ibn al-Bannā de Marrakech, sufi et mathématicien (XII^{e}–XIII^{e} s. J. C.), *Hesperis* 25 (1938), 33.
31. Ibn Khaldūn (see Note 7, p. 897).
32. Ibn Munᶜim refers here to al-Khalīl's *Kitāb al-ᶜAyn* (The Book of ᶜAyn), and more precisely to the passage on the counting of words of the Arabic language composed of two, three, four, or five phonemes.
33. This is, moreover, the same approach that Mersenne followed in the 17th century before Frénicle established the formula giving what he called the 'combinations of changes' in which 'several similar things are formed'. See E. Coumet, *Mersenne, Frénicle . . .*, pp. 248–61, and B. Frénicle, *Abrégé des combinaisons*, in *Mémoires de l'Académie Royale des Sciences*, 1666–99, t. V, Paris (1729), 99–105.
34. Ibn al-Bannā, *Tanbīh al-albāb* (Warning to Intelligent [People]), Ms. Alger. B. N. 613/6, f. 73a–b.
35. Ibn al-Bannā, *Rafᶜ al-ḥijab ᶜan wujūh aᶜmāl al-ḥisāb* (Lifting of the Veil on the Different Operations of Calculation), Tunis B. N., Ms. 9722, ff. 12a; see also M. Aballagh, *Rafᶜ al-ḥijab d'Ibn al-Bannā*, Doctoral thesis, University of Paris I–Pantheon–Sorbonne (1988).
36. A. Djebbar (see Note 10).
37. Ibn al-Bannā (see Note 35, ff. 42a,b).
38. Ibn al-Bannā (see Note 35, f. 5a and ff. 8b–9a).
39. Ibn al-Bannā, *Risāla fī t-taksīr* (Epistle on Measuring), Tunis B. N., Ms. 9002, ff. 130a–132b.
40. Ibn al-Bannā, *Arbaᶜ maqālāt* (The Four Epistles), Tunis B. N., Ms. 9722, ff. 116a–121a.
41. Ibn al-Bannā (see Note 35, f. 42b, where the author revisited the result already established with the letters of the alphabet), whereas Ibn Haydūr used the correspondence between n objects and a subset of $\{1, 2, \ldots, n\}$ in his *al-Jāmiᶜ fī l-ḥisāb* (Collections of Calculations), Tunis B. N., Ms. 9722, f. 49b.
42. Ibn al-Bannā (see Note 34, ff. 69a–74a). This question concerns the following phrase: *ḍaraba aḍ-ḍāribu ash-shātimu al-qātilu muḥibbaka widdaka qāṣidaka muᶜjaban khālidan*. Each word of this phrase can, without changing place, play a certain number of grammatical functions (seven for the 7th word of the phrase, six for the 6th, four for the 5th, six for the 4th, six for the 3rd, five for the 8th, and nine for the last): this gives 272 160 different readings of the same phrase.

43. Ibn Haydūr counted types of fractions and systems of equations. Ibn al-Majdī revisited the question of counting certain types of fractions, and also interested himself in the problem of prayers which was posed and resolved by Ibn al-Bannā in his *Tanbīh al-albāb*, where he gave a generalization (see A. Djebbar, Note 10).
44. Ibn Khaldūn (see Note 7, pp. 1059–60).
45. See al-Bīrūnī, *Kitāb Maqālīd ᶜilm al-hay'a* (Book of the Keys to Astronomy) and M. V. Villuendas (ed.), *La Trigonometria Europea en el Siglo XI, Estudio de la Obra de Ibn Muᶜād El Kitāb Mayhūlāt*, Instituto de Historia de la Ciencia de la Real Academia, Barcelona (1979).

Title page of the *Sefer Yetsirah*.

CHAPTER 4

Jewish combinatorics

VICTOR J. KATZ

Many Jewish scholars from the early years of our era were interested in calculating permutations and combinations. Among the problems that led to the study of these notions were finding the number of words that could be formed out of the letters of the Hebrew alphabet, and determining the number of conjunctions of the planets. It was Levi ben Gerson in the 14th century who was able to formalize these notions and rigorously derive the formulas for the numbers of permutations and combinations.

Introduction

Although Jews today are relatively active in the mathematical sciences, restrictions on Jewish life in Europe in the medieval period made it difficult for Jews at that time to study mathematics. Nevertheless, at various times and places in medieval Europe, when Jews were allowed to participate in public life, some Jews did make contributions to mathematical ideas. In particular, Jews were active in intellectual life in medieval Spain under Muslim rule, under the Christians in Provence before it became part of France in the 15th century, and in various Italian states at the time. The mathematical ideas in which Jews were interested generally stemmed from questions arising from their Biblical and Talmudic studies, since any Jewish man who showed scholarly bent began his studies with religious texts. Therefore, invariably, Jewish scholars of the time were rabbis, because it was rabbis who would be supported by the Jewish community at large. And it was their rabbinical training that often led them to go beyond answers to immediate practical questions and consider more abstract and theoretical results.

In this chapter we consider Jewish contributions to the study of combinatorics – specifically, to the study of permutations and combinations. The first analysis of such questions arose from specific matters: thus, the study of permutations arose from the question of how many words can be created from the letters of the Hebrew alphabet. The question was a mystical one: God created the world by giving names to things, so the rabbis wanted to know how many 'things' God could have named. The first appearance of this question seems to be in the *Sefer Yetsirah* (see below), but it continued to appear in the works of authors as late as the 16th century.

Similarly, the question of combinations showed up initially as an astrological question in the work of Rabbi Abraham ibn Ezra in the 12th century. Ibn Ezra wanted to know how many possible conjunctions there are of the seven planets, for a conjunction was a significant astronomical event, possibly portending major consequences on earth.

Eventually, Rabbi Levi ben Gerson in the 14th century was able to treat both permutations and combinations as purely mathematical questions. In a major work of 1321 Levi worked out, often using the method of mathematical induction, the formulas for determining the number of permutations and combinations in the case of arbitrary finite sets.

The *Sefer Yetsirah*

The earliest surviving Jewish source on combinatorics seems to be the mystical work *Sefer Yetsirah* (Book of Creation), written sometime before the 8th century and perhaps as early as the 2nd century. In this work, the unknown author calculated the various ways in which the twenty-two letters of the Hebrew alphabet can be arranged. Jewish mystics believed that God had created the world and everything in it by naming these things (in Hebrew, of course), so it was of interest to know how many 'things' could be named (see [4, p. 23]):

> God drew them, combined them, weighed them, interchanged them, and through them produced the whole creation and everything that is destined to be created.

א ב ג ד ה ו ז ח
ט י כ ל מ נ ס ע
פ צ ק ר ש ת

The twenty-two letters of the Hebrew alphabet.

In one passage the author wrote about making words from just two letters (see [4, p. 11]):

He fixed the twenty-two letters on the sphere like a wall with two hundred and thirty-one gates, and the sphere turns forwards and backwards . . . But how was it done? He combined . . . the aleph with all the other letters in succession, and all the others again with aleph; bet with all and all again with bet, and so the whole series of letters. Hence it follows that there are two hundred and thirty-one formations.

What this seems to mean is that the author found that there are 231 'forward' combinations of two letters, where the earlier letter comes before the later one, and then 231 'backward' ones, where the later letter comes before the earlier one. The 10th-century commentator Saadia Gaon (892–942), who was born in Egypt but spent much of his life in Babylonia as head of a Talmudic academy, noted that children in Palestine learned spelling and pronunciation by considering all $22^2 = 484$ possible ordered pairs of letters. He noted further that the author of the *Sefer Yetsirah* dropped all twenty-two repeated pairs from consideration and then halved the remaining number to get his result.

The author also considered how to make words from more than two letters (see [4, p. 23]):

Two stones [letters] build two houses [words], three build six houses, four build twenty-four houses, five build one hundred and twenty houses, six build seven hundred and twenty houses, seven build five thousand and forty houses. From thence further go and reckon what the mouth cannot express and the ear cannot hear.

Evidently, the author understood that the number of possible arrangements of n letters is $n!$.

Saadia Gaon extended this rule as follows (see [1, p. 495]):

> If somebody wants to know how many words may be built from a larger number than that, as for instance 8, 9, 10 and so on, the rule is that one should multiply the result of the first product by the following number and what one thus obtains is the sum total. And its explanation is as follows: the permutations of two letters give 2 words, if you multiply 2 by 3, you get 6, and that is the number of the permutations of the three letters . . . If you want to know the number of the permutations of 8 letters, multiply the 5040 that you got from 7 by 8 and you will get 40 320 words; and if you search for the number of permutations of 9 letters, multiply 40 320 by 9 and you will get 362 880; and if you search for the number of permutations of 10 letters, multiply 362 880 by 10 and you will get 3 628 800 words; and if you search for the number of permutations of 11, multiply these 3 628 800 by 11 and you will get 39 916 800 words. And if you want to know [the permutations] of still larger numbers, you may operate according to the same method. We, however, stopped at the number of 11 letters, for the longest word to be found in the Bible [with no letter repeated] contains 11 letters.

Evidently, the author understood that $n! = n \times (n-1)!$.

An Italian rabbi, Shabbetai Donnolo (913–70), explicitly derived the factorial rule for permutations in a somewhat later commentary on the *Sefer Yetsirah* (see [6, p. 144]):

> The first letter of a two-letter word can be interchanged twice, and for each initial letter of a three-letter word the other letters can be interchanged to form two two-letter words – for each of three times. And all the arrangements there are of three-letter words correspond to each one of the four letters that can be placed first in a four-letter word: a three-letter word can be formed in six ways, and so for every initial letter of a four-letter word there are six ways – altogether making twenty-four words, and so on.

Since the *Sefer Yetsirah* continued to be studied in Jewish mystical circles, we even find much later commentaries on the mathematics involved. In the 16th century, for example, Rabbi Moses Cordovero (1522–70) from Safed, Palestine, gave a very detailed explanation of the basic permutation rule, and then proceeded to generalize his results to the case where one or more letters in a word are duplicated (see [7, p. 32]):

> If we should find a repeated letter, then one half of the number corresponding to the number of elements should be subtracted. For this reason the tetragrammaton [the four letters making up the Hebrew name of God, two of which are the same] will not give more than 12 constructions, or one half the number otherwise resulting – since four stones usually build 24 houses, while these four stones build but 12 . . . A word of three letters

would by the law give rise to six forms; if, however, there are two like letters in the group there will be but three permutations. If, in a group of four, a letter be repeated three times the first duplication will destroy half of the permutations, as we have explained; and the second duplication (that is, the third like letter) will destroy two-thirds of the remainder, four forms remaining out of 12. Thus *abbb*, which should give 24 permutations by the original law, will (because of three like letters) yield but four. In a group of five letters whose permutations would ordinarily be 120, if there are two like letters as in *abcdd*, there will remain 60; if there are three like letters, as in *abddd*, two-thirds of the remaining forms will disappear and but 20, or one-third, will remain as a result of the effect of the third letter . . .

Cordovero completed his discussion by considering the case where more than one letter is repeated.

Abraham ibn Ezra

As noted earlier, the author of the *Sefer Yetsirah* briefly mentioned how to calculate the number of combinations of letters taken two at a time. A more detailed study of combinations was carried out by Rabbi Abraham ben Meir ibn Ezra (1090–1167), a Spanish–Jewish philosopher, astrologer, and Biblical commentator. He left his native Spain in 1140 and spent the remainder of his life in various Jewish communities in Italy and the south of France, at which time he composed most of his Biblical commentaries.

Ibn Ezra discussed combinations in an astrological text. In particular, because conjunctions of the seven planets (including the Sun and the Moon) were believed to have a powerful influence on human life, he wanted to count the ways in which these could occur. Ibn Ezra in fact calculated $C(7, k)$ for each integer k from 2 to 7 and noted that the total was 120. He began with the simplest case, that the number of binary conjunctions (sets of two elements out of seven) is 21. This number is equal to the sum of the integers from 1 to 6, and can be calculated by a well-known rule. Ibn Ezra wrote the calculation this way (see [2, p. 350] and [3]):

It is known that there are seven planets. Now Jupiter has six conjunctions with the [other] planets. Let us multiply then 6 by its half and by half of unity. The result is 21, and this is the number of binary conjunctions.

Although ibn Ezra left it to the reader to understand that Saturn has five conjunctions, Mars four, and so on, what he was asserting, in modern terms, is that

$$C(7,2) = 1+2+3+4+5+6 = 6 \times \left((\tfrac{1}{2} \times 6) + \tfrac{1}{2}\right) = 21.$$

More generally, ibn Ezra asserted that

It is known that the sum of the numbers from 1 to any desired number is found by multiplying it by half of itself and by half of unity.

That is,

$$1+2+\cdots+n = n \times \left((\tfrac{1}{2} \times n) + \tfrac{1}{2}\right) = \tfrac{1}{2}n(n+1).$$

To calculate ternary combinations, ibn Ezra first calculated the number of these involving Jupiter (see [2, p. 351]):

We begin by putting Saturn with Jupiter and with them one of the others. The number of the others is five; multiply 5 by its half and by half of unity. The result is 15. And these are the conjunctions of Jupiter.

Thus, there are five ternary combinations involving Jupiter and Saturn, four involving Jupiter and Mars, but not Saturn, and so on. Hence, there are

$$C(6,2) = 5 \times \left((\tfrac{1}{2} \times 5) + \tfrac{1}{2}\right) = 15$$

ternary conjunctions involving Jupiter.

Similarly, to find the ternary conjunctions involving Saturn but not Jupiter, ibn Ezra needed to calculate the number of choices of two planets from the remaining five: $C(5,2) = 10$. He then found the ternary conjunctions involving Mars, but neither Jupiter nor Saturn, and concluded with the result

$$\begin{aligned} C(7,3) &= C(6,2) + C(5,2) + C(4,2) + C(3,2) + C(2,2) \\ &= 15 + 10 + 6 + 3 + 1 = 35. \end{aligned}$$

He next calculated the quaternary conjunctions by analogous methods. The conjunctions involving Jupiter require choosing three planets from the remaining six; those with Saturn but not Jupiter require choosing three from five. So, finally,

$$C(7,4) = C(6,3) + C(5,3) + C(4,3) + C(3,3) = 20 + 10 + 4 + 1 = 35.$$

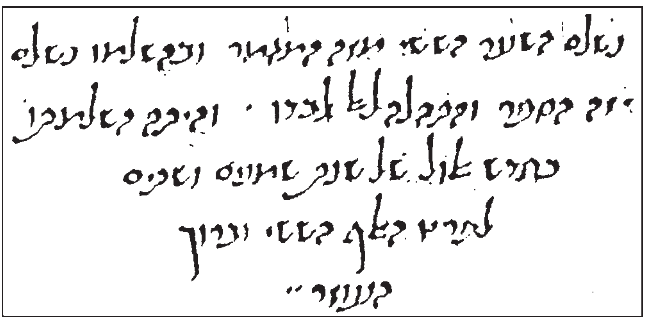
Ibn Ezra simply stated the results for the conjunctions involving five, six, and seven planets. Essentially, his work gives an argument for the case $n = 7$, easily generalizable to the general combinatorial rule

$$C(n,k) = C(n-1,k-1) + C(n-2,k-1) + \cdots + C(k-1,k-1).$$

Rabbi ibn Ezra's astrological work, which included this material on combinations, appeared in Latin translation in 1281. The Latin version seems to indicate that the translation was made from the Arabic, so it must sometime earlier have been translated into Arabic from the Hebrew original, although no such Arabic manuscript has yet been found. Nevertheless, some of these same ideas appear in Arabic writings of the 13th century.

Levi ben Gerson

Levi ben Gerson (1288–1344) was an astronomer, philosopher, mathematician, and Biblical commentator, who spent his entire life in and around Orange, in the south of France. Although today he is known in Jewish circles mainly for his philosophical works, in his time he was well regarded even in the Christian scientific community of France for his knowledge of mathematics and astronomy. In 1321, he wrote his most important mathematical work, the *Maasei Hoshev* (The Art of the Calculator), in which he gave careful rigorous proofs of various combinatorial formulas.



An extract from Levi ben Gerson's *Maasei Hoshev*.
Translation: 'The sixth section of this volume is finished, and with its completion the book is complete. The praise goes exclusively to God. Its completion was in the month of Elul in the 82nd year of the sixth millennium. Bless the Helper.'

Levi's text, which appeared in a second edition in 1322, is divided into two parts – a first theoretical part, in which each result receives a detailed proof, and a second applied part, in which explicit instructions are given for performing various types of calculation and which contains numerous interesting problems, most to be solved using ratio and proportion. Levi's theoretical first section begins with a quite modern justification for considering theory at all (see [5, p. 1]):

Because the true perfection of a practical occupation consists not only in knowing the actual performance of the occupation but also in its explanation, why the work is done in a particular way, and because the art of calculating is a practical occupation, it is clear that it is pertinent to concern oneself with its theory. There is also a second reason to inquire about the theory in this field. Namely, it is clear that this field contains many types of operations, and each type itself concerns so many different types of material that one could believe that they cannot all belong to the same subject. Therefore, it is only with the greatest difficulty that one can achieve understanding of the art of calculating, if one does not know the theory. With the knowledge of the theory, however, complete mastery is easy. One who knows it will understand how to apply it in the various cases which depend on the same foundation. If one is ignorant of the theory, one must learn each kind of calculation separately, even if two are really one and the same.

Levi worked up gradually to the formulas for permutations and combinations, at each step giving a careful Euclidean-style proof. In some of these proofs, he used the essentials of the method of mathematical induction, which he called the process of 'rising step by step without end'. In general, when Levi used such a proof, he first proved the *inductive step*, the step that allows one to move from k to $k+1$, then noted that the process begins at some small value of k, and finally stated the complete result. However, given that he had no notation for an arbitrary integer (which we indicate by k), what he always did in this context was prove the inductive step by the method of generalizable example. In other words, he proved this step for a particular integer, making it clear that the proof would be the same no matter what integer was picked. (This method of generalizable example was used frequently in pre-modern mathematics, by many mathematicians from Euclid to Pascal.)

Of course, Levi did not give a general statement of his inductive proof method. On the other hand, he used it in various proofs throughout the book, including two of the earliest theorems in the text, theorems that deal with associativity and commutativity of multiplication.

Proposition 9. If one multiplies a number which is the product of two numbers by a third number, the result is the same as when one multiplies the product of any two of these three numbers by the third.

Proposition 10. If one multiplies a number which is the product of three numbers by a fourth number, the result is the same as when one multiplies the product of any three of these four numbers by the fourth.

In modern notation, the first result states that $a(bc) = b(ac) = c(ab)$, while the second result extends this result to four factors.

Levi proved Proposition 9 by counting the number of times that the various factors of the product appear in that product. In his proof of Proposition 10 he noted that $a(bcd)$ contains bcd a times. Since, by Proposition 9, bcd can be thought of as $b(cd)$, it follows that the product $a(bcd)$ contains acd b times, so $a(bcd) = b(acd)$, as desired. Levi then generalized these two results to any number of factors (see [5, p. 8]):

> By the process of rising step by step without end, this is proved; that is, if one multiplies a number which is the product of four numbers by a fifth number, the result is the same as when one multiplies the product of any four of these by the other number. Therefore, the result of multiplying any product of numbers by another number contains any of these numbers as many times as the product of the others.

We see here the essence of the principle of mathematical induction, in that both steps of the method are explicitly proved, with the second step proved by the method of generalizable example. Levi used the principle again in proving that $(abc)d = (ab)(cd)$, and concluded that one can use the same proof to demonstrate the following result:

> Any number contains the product of two of its factors as many times as the product of the remaining factors.

Levi was certainly not consistent about applying his induction principle. The middle of the text contains many theorems that deal with sums of various sequences of integers – theorems that could be proved by induction – but for many of these Levi used other methods. In his proof of the formula for the sum of the first n integral cubes, however, he used a method related to induction that we would call the 'method of descent'. The critical result here is, in fact, the result we would use today in giving an inductive proof of this formula:

Proposition 41. The square of the sum of the natural numbers from 1 *up to a given number is equal to the cube of the given number added to the square of the sum of the natural numbers from* 1 *up to one less than the given number.*

In modern notation, this result says that

$$(1 + 2 + \cdots + n)^2 = n^3 + (1 + 2 + \cdots + (n - 1))^2.$$

We present Levi's proof symbolically. First, by Levi's Proposition 30,

$$n^2 = (1 + 2 + \cdots + n) + (1 + 2 + \cdots + (n - 1)).$$

Then

$$n^3 = n \times n^2 = n\Big((1 + 2 + \cdots + n) + (1 + 2 + \cdots + (n - 1))\Big)$$
$$= n^2 + n\Big(2 \times (1 + 2 + \cdots + (n - 1))\Big).$$

But

$$(1 + 2 + \cdots + n)^2 = n^2 + 2n(1 + 2 + \cdots + (n - 1)) + (1 + 2 + \cdots + (n - 1))^2.$$

It follows that

$$n^3 + (1 + 2 + \cdots + (n - 1))^2 = (1 + 2 + \cdots + n)^2.$$

Levi next noted that, although 1 has no number preceding it, 'its third power is the square of the sum of the natural numbers up to it'. In other words, he gave the first step of a proof by induction for the result stated as follows:

Proposition 42. The square of the sum of the natural numbers from 1 *up to a given number is equal to the sum of the cubes of the numbers from* 1 *up to the given number.*

Levi noted that, by Proposition 41,

$$(1 + 2 + \cdots + n)^2 = n^3 + (1 + 2 + \cdots + (n - 1))^2.$$

The final summand is, also by Proposition 41, equal to

$$(n - 1)^3 + (1 + 2 + \cdots + (n - 2))^2.$$

Continuing in this way, Levi eventually reached $1^2 = 1^3$, and the result is proved.

We note again that, although the proposition is stated in terms of an arbitrary natural number, Levi's own proof is by the method of generalizable example. In this case his proof was given for $n = 5$; in fact, the first five numbers were represented by the five initial letters of the Hebrew alphabet.

Levi did use mathematical induction to prove the major results on permutations and combinations, the theorems that provided the climax to the theoretical part of the text. First, he showed that the number of permutations of a given number n of elements is what we now call $n!$:

Proposition 63. If the number of permutations of a given number of different elements is equal to a given number, then the number of permutations of a set of different elements containing one more number equals the product of the former number of permutations and the given next number.

Symbolically, this proposition states that

$$P(n+1) = (n+1)P(n),$$

where $P(n)$ stands for the number of permutations of a set of n elements. This result provides the inductive step in the proof of the proposition $P(n) = n!$, although Levi did not mention that result until the end.

His proof of Proposition 63 was very detailed. Given a permutation *abcde* of the original n elements, and a new element f, he noted that *fabcde* is a permutation of the new set. Because there are $P(n)$ such permutations of the original set, there are also $P(n)$ permutations of the new set beginning with f. Also, if one of the original elements, such as e, is replaced by the new element f, then there are $P(n)$ permutations of $\{a, b, c, d, f\}$ and therefore also $P(n)$ permutations of the new set with e in the first place. Because any of the n elements of the original set, as well as the new element, can be put in the first place, it follows that the number of permutations of the new set is $(n+1)P(n)$. Levi finished the proof of Proposition 63 by showing that these $(n+1)P(n)$ permutations are all different. He concluded (see [5, p. 49]):

> Thus it is proved that the number of permutations of a given set of elements is equal to that number formed by multiplying together the natural numbers from 1 up to the number of given elements. For the number of permutations of 2 elements is 2, and that is equal to $1 \cdot 2$, the number of permutations of 3 elements is equal to the product $3 \cdot 2$, which is equal to $1 \cdot 2 \cdot 3$, and so one shows this result further without end.

Thus, Levi mentioned the beginning step and then noted that, with the inductive step already proved, the complete result is also proved.

After proving by a counting argument that $P(n, 2) = n \times (n-1)$, where $P(n, k)$ represents the number of permutations of k elements in a set of n elements, Levi proved by induction on k that

$$P(n, k) = n \times (n-1) \times (n-2) \times \cdots \times (n-k+1).$$

As before, he stated the inductive step as a theorem:

Proposition 65. If a certain number of elements is given and the number of permutations of order a number different from and less than the given number of elements is a third number, then the number of permutations of order one more in this given set of elements is equal to the number which is the product of the third number and the difference between the first and the second numbers.

Modern symbolism replaces Levi's convoluted wording with a brief phrase:

$$P(n, k+1) = (n-k) \times P(n, k).$$

Levi's proof was quite similar to that of Proposition 63, in that he looked at how permutations of $k+1$ elements can be found from those of k elements. At the end, he stated the complete result (see [5, p. 51]):

> It has thus been proved that the permutations of a given order in a given number of elements are equal to that number formed by multiplying together the number of integers in their natural sequence equal to the given order and ending with the number of elements in the set.

To clarify this statement, Levi again used a generalizable example. He gave the initial step of the induction by quoting his previous result in the case $n = 7$ – that is, the number of permutations of two elements in a set of seven is equal to 6×7. Then the number of permutations of three elements is equal to $5 \times 6 \times 7$ (since $5 = 7 - 2$). Similarly, the number of permutations of four elements is equal to $4 \times 5 \times 6 \times 7$, 'and so one proves this for any number'.

In the final three propositions of the theoretical part of *Maasei Hoshev*, Levi completed his development of formulas for permutations and combinations. Proposition 66 shows that

$$P(n, k) = C(n, k) \times P(k, k),$$

while Proposition 67 simply rewrites this as

$$C(n, k) = P(n, k)/P(k, k).$$

Since he had already given formulaic procedures for calculating both the numerator and the denominator of this quotient, Levi had thus demonstrated the equivalent of the standard formula for $C(n, k)$:

$$C(n, k) = \frac{n \times (n-1) \times (n-2) \times \cdots \times (n-k+1)}{1 \times 2 \times \cdots \times k}.$$

Finally, Proposition 68 demonstrates that $C(n, k) = C(n, n-k)$.

Thus, by 1321, the basic results on permutations and combinations were available in the Principality of Orange. Unfortunately for the development of mathematics, Levi's work does not seem to have had any influence in subsequent centuries. So, did anyone read Levi's book? Given that there are today about a dozen manuscript copies of the work extant, in libraries throughout Europe as well as one in New York, it appears that the book must have had a reasonable circulation for a medieval manuscript. In particular, we know that there was a copy in Paris in the library of the Oratorian priests, a copy brought there from Constantinople in 1620 by Achille Harlay de Sancy, the French ambassador to the Ottoman Empire. This copy would certainly have been available to any of the priests at the Oratory who read Hebrew and were trained in mathematics, and possibly to Marin Mersenne (1588–1648), who was in contact with many of these priests and who included a study of combinations and permutations in his work on music theory in the 1630s. But Mersenne made no reference to this text and only some cryptic remarks about how he learned combinatorics himself. And even though Blaise Pascal (1623–62) certainly learned some mathematics in his visits to Mersenne's gatherings of mathematicians, he too made no reference to previous writers on the theory of induction in his own statements on the subject. Thus, as is true for many results in the history of mathematics, the original discoverers received no mention when their ideas were rediscovered years later and finally brought into the mathematical mainstream.

References

1. S. Gandz, Saadya Gaon as a mathematician, *American Academy of Jewish Research Texts and Studies*, Vol. II (1943); also found in S. Gandz, *Studies in Hebrew Astronomy and Mathematics*, Ktav Publishing House (1970).
2. J. Ginsburg, Rabbi ben Ezra on permutations and combinations, *Math. Teacher* 15 (1922), 347–56.
3. A. ibn Ezra, *Ha-Olam* (English transl. in [2]; Latin transl. in M. Clagett's *Nicole Oresme and the Medieval Geometry of Qualities and Motions*, University of Wisconsin Press (1968), 445–6).
4. I. Karlisch (ed. and transl.), *The Sepher Yezirah* (The Book of Formation), Heptangle Books (1987).
5. G. Lange, *Sefer Maasei Choscheb. Die Praxis der Rechners. Ein hebraisch-arithmetisches Werk des Levi ben Gerschom aus dem Jahre 1321*, Louis Golde (1909).
6. N. L. Rabinovitch, *Probability and Statistical Inference in Ancient and Medieval Jewish Literature*, University of Toronto Press (1973).
7. M. Turetzky, Permutations in the 16th century Cabala, *Math. Teacher* 16 (1923), 29–34.

HARMONIE VNIVERSELLE.

Ex antiquo marmore Illustrissimi Marchionis Mathei Romæ.

Nam & ego confitebor tibi in vasis psalmi veritatē tuam: Deus psallam tibi in Cithara, sanctus Israel. *Psalme 70.*

Title page of Marin Mersenne's *Harmonie Universelle*.

CHAPTER 5

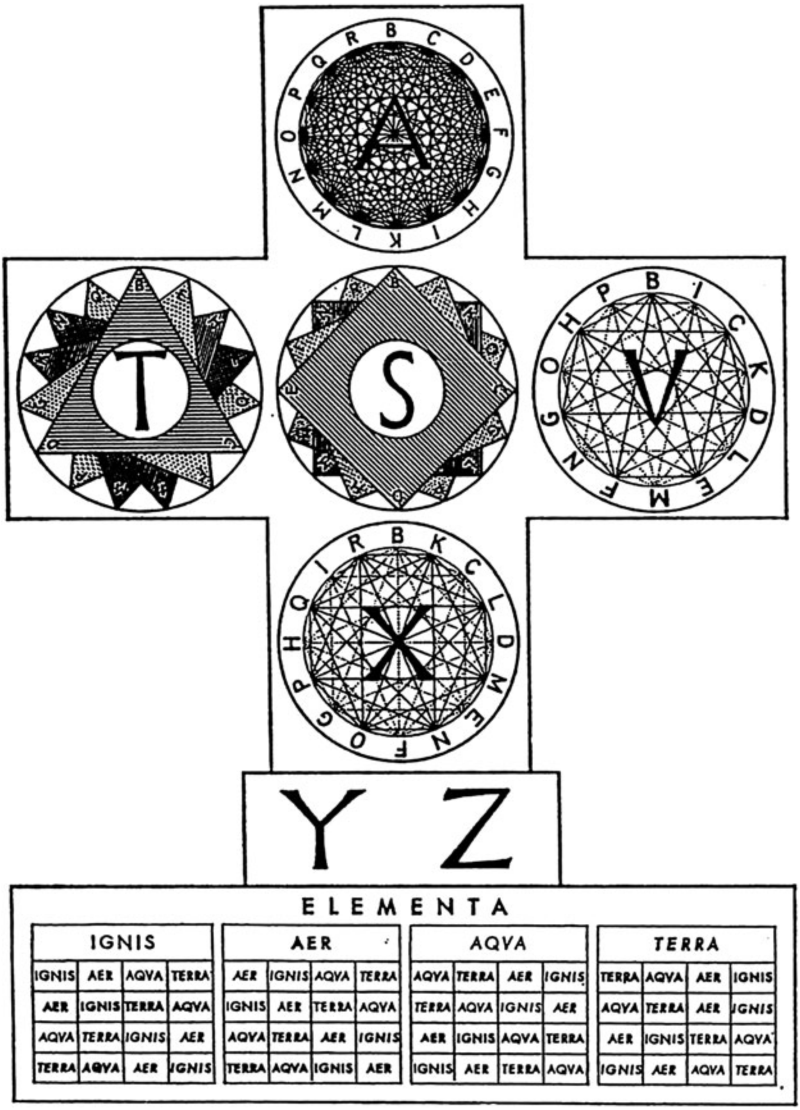
Renaissance combinatorics

EBERHARD KNOBLOCH

Combinatorial thinking in the Renaissance had philosophical, religious, and game-theoretical roots. The concept of Llullism, in which all knowledge is derived by combining a finite number of attributes, originated in the 13th century and spread throughout Europe; in the same century we find combinatorial studies related to games of dice. From the 16th century, Jesuits, Cistercians, and members of other religious orders played a crucial role in the development of combinatorics. The basic combinatorial operations were explained and illustrated by examples from daily life and by tables, normally without proof: number theory and music theory were the most important mathematical fields of application. In the 17th century, authors frequently inserted sections on combinatorics into their arithmetic or algebra textbooks, and began to write special monographs on the subject.

Medieval glimpses

One of the two main roots of combinatorial thinking in the Middle Ages was philosophical in nature. The Catalan mystic, poet, and missionary Ramon Llull (1232–1316) was by far the most influential author in this regard. His 'Great or General Art' was a systematic summary of all the branches of knowledge of his time, based on the combinatorial art. In order to know something it was not necessary to analyse the objects by means of experience, but rather to understand the fundamental terms, the so-called ‘principles’, and to combine them with each other. Thus, combinatorics became the basic tool for exploring all that was known at that time. The Llull school remained important for several hundred years, and flourished especially in the 17th century (see [29]).



Some of Ramon Llull’s circulatory diagrams.

Llull illustrated his method by mechanisms of letters, figures, triangles, and circles, and introduced six symmetric figures:

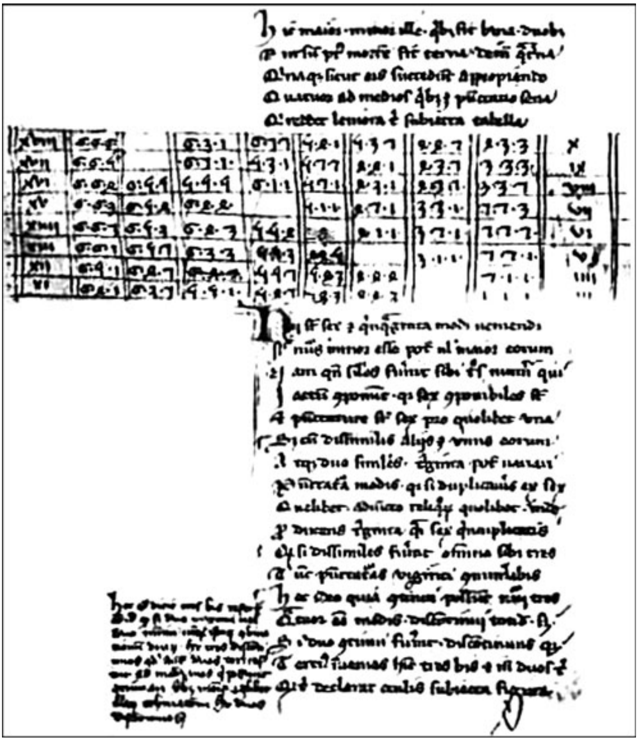
- figure A depicts the sixteen fundamental values (*dignitates*), such as goodness, magnitude, power, wisdom, and truth;
- figure S contains twelve composed notions and four combinatorial figures;
- figure T concerns relative principles and significations;
- figure V links the virtues and vices;
- figure X concerns the predestination of individuals;
- figure YZ combined the figures A, S, T, U, V, and X, in the form of a crucifix.

Underneath are four boxes, whose contents (again, the four elements) are ordered as latin squares; in the last two columns, the second entries should be IGNIS and AER.

Unlike his followers in the 17th century, Llull was not primarily interested in the mathematical problem of enumerating the possible sign combinations. His combinatorial art was designed as an inventive logic that could judge every posed problem. Yet, Llull drew attention to such mathematical calculations.

The second main source was games of chance – especially dice games. The pseudo-Ovidian hexametric poem *De Vetula* (On a Little Old Woman) was written between 1222 and 1268, probably by Richard de Fournival (b.1201). There the author enumerated the fifty-six essentially different throws of three dice: 'essentially different' means that the points of the dice are not counted individually – only the types of combination of the points matter. He correctly explained why there are thirty cases where two dice show equal points, and twenty cases where all three dice have distinct points. He added the numbers under the condition that the dice are individualized, and he knew that the sum of these numbers is 216 (see [17] and [31]).

In 1283 the Castilian king Alphons X the Wise (1221–84) composed his *Libros de Acedrez, Dados e Tablas* (Books on Chess, Dice, and Tables). The section on games of dice describes twelve such games and enumerates the fifty-six possible throws of three dice generating from 3 up to 18 points, and the possible throws of two dice for which the sum of their values equals the number on the third die. Similar considerations for three dice appear in a commentary on Dante's *Divina Commedia* (Divine Comedy), published in 1477.



An extract from *De Vetula*.

Italian beginnings

In 1494 Leonardo da Vinci's friend Luca Pacioli (1445–1517) published his mathematical encyclopedia, *Summa de Arithmetica Geometria Proportioni e Proportionalità* (Comprehensive Treatise on Arithmetic, Geometry, Proportions, and Proportionality). Here, he studied the arrangements of a number of people at a table (2nd distinction, 5th treatise), calculating 11! or $P(11, 11) = 39\,916\,800$; to this end he used the recursion law

$$n \times (n-1)! = n!$$

(see [28] and [19, pp. 2f]).

Much more combinatorial knowledge was revealed by Girolamo Cardano, one of the most important scientists of the 16th century. He was a successful physician who wrote on many subjects, and his own list of works includes forty-six monographs. His *Ars Magna* (The Great Art) on algebra (1545) explained the algorithmic solution of the general cubic equation, which he had learned from his countryman, the arithmetician and physicist Niccolò Tartaglia (*c.*1500–57). At least four of his books deal with combinatorics: his *Practica Arithmeticae* (Practice of Arithmetic) of 1539 [6], *De Subtilitate* (On Subtlety) of 1550 [7], *Liber de Ludo Aleae* (Book on the Game of Dice) written around 1550 [9], and his *Opus Novum de Proportionibus* (New Work on Proportions) of 1570 [8].

Girolamo Cardano (1501–76).

In Chapter 51 of his *Practica Arithmeticae*, Cardano gave several examples to illustrate the number $2^n - 1 - n$ of all combinations of at least two things selected from n: eleven different powers of x admit $2^{11} - 1 - 11 = 2036$ linear combinations, the seven planets admit $2^7 - 1 - 7 = 120$ combinations, and twenty-two letters lead to $2^{22} - 1 - 22 = 4\,194\,281$ expressions (*dictiones*). Their difference, he said, concerns the substance, not the order. Still more expressions occur if the order is permuted; thus Cardano hinted also at arrangements. It is worth noting that, for Cardano, a combination involved at least two elements: for him, selections of one or no elements were not combinations.

In Chapter 15 of *De Subtilitate*, Cardano explained how the numbers $C(n,k) = C(n, n-k)$ could be calculated from the preceding values $C(n, k-1) = C(n, n-k+1)$ by multiplying by $n-k+1$ and dividing by k. He took both selections into account at the same time – for example, he wrote, for $n = 20$,

$$(40 \times 19)/2 = 2 \times (20 \times 19)/2 = C(20, 2) + C(20, 18),$$

and continued with $(40 \times 19 \times 18)/(2 \times 3)$, etc. Thus, already by 1550, Cardano knew the multiplicative calculation of $C(n,k)$:

$$C(n,k) = \frac{n \times (n-1) \times \cdots \times (n-k+1)}{1 \times 2 \times \cdots \times k}.$$

In Chapter 17, he explained the construction of 'combination locks' consisting of several rotating numbered rings, but he did not analyse the related combinations of numbers mathematically: such analyses are to be found in his *Liber de Ludo Aleae*. This book did not influence the development of the subject because it was not published until 1663, in an edition of his complete works.

In order to study dice probabilities Cardano intensively investigated the possible throws of two or three dice. He calculated the number of arrangements with repetition in special cases, and found the number of throws or combinations that meet certain conditions: for example, there are eleven throws of two dice that include at least one 1, and ninety-one ways of obtaining at least one 6 in three throws of a die, or in one throw of three dice. He also classified the throws of three dice according to the type of repetition: for example, there are six throws with all three faces alike, ninety throws with a double and one single, such as $(1,1,2)$, and 120 throws with all faces different, such as $(1,2,3)$.

In a similar way he solved the question of finding how many ways a throw of two dice, like $(1,2)$, can be completed to a throw of three dice; the result looked for is 30 – that is, there are thirty throws of three dice that include at least one 1 and at least one 2. Finally, he considered the different types of throws of four astragals (heel bones in the shape of three-sided prisms with rounded ends): there are thirty-five such types of throws (as explained in [9], p. 276).

In order to solve such problems, Cardano considered specific permutations with and without repetition, but without mentioning any general rules for calculating these numbers; for example, since there are twenty throws where all faces are different, and each one gives rise to six different permutations, there are $6 \times 20 = 120$ such throws altogether.

All in all, Cardano's dice results represented a notable achievement, and far more than Galileo Galilei achieved in his fragment *Sopra le Scoperte dei Dadi* (On Findings of Dice). Galileo probably wrote this between 1613 and 1623 in order to explain the different frequencies of throwing the totals 9 and 10 with three dice (see [27]).

In his *Opus Novum de Proportionibus*, Cardano resumed his combinatorial studies. With the aid of numerical examples he explained (in words) the additive law of formation

$$C(n,k) + C(n,k+1) = C(n+1,k+1),$$

used the multiplicative law in the form

$$C(n,k) = \frac{n-(k-1)}{k} \times C(n,k-1)\,,$$

and illustrated the equality

$$C(n,k) = C(n,n-k)$$

by means of the arithmetical triangle. This time he determined the number of all possible combinations of n elements as $2^n - 1$, since he now included selections with just one element among his combinations.

Meanwhile, the arithmetician and physicist Niccolò Tartaglia (1499/1500–57) had published his *General Trattato di Numeri et Misure* (General Treatise on Numbers and Measures) in 1556 (see [38] and [19, p. 3]). He repeated Pacioli's rule for permutations without repetition, and determined the number of special combinations with repetition by enquiring into the number of essentially different throws of up to eight dice. To this end he relied on the arithmetical triangle and used the identity we now write as

$$\sum_k C(n+k-1,k) = C(n+k,k).$$

Outside Italy

The combinatorial results of the Italian mathematicians were quickly taken up by authors from outside Italy. The German parson and algebraist Michael Stifel (*c.*1487–1567) explicitly referred to 'Cardano's rule' in his *Arithmetica Integra* (Complete Arithmetic), regarding the ways of combining arbitrary given things (see [35]): 'It is useful', he said, 'to find the number of aliquot parts of arbitrary numbers'. Stifel alluded to Cardano's *Practica Arithmeticae* and to the $(2^n - 1 - n)$-rule, modifying it to $2^n - 1$ in order to find the number of proper factors ('aliquot parts') of a product of n distinct prime numbers. The same distinction between the number of combinations of elements and the number of proper factors of a product was also made by William Buckley in his *Arithmetica Memorativa* [Arithmetic Useful for Memorizing], which appeared in 1567 (see [2] and [12]).

Italian influence was also apparent in the *Logistica* (Arithmetic) of the French mathematician Jean Borrel (*c.*1492–*c.*1570) (see [1]). By 1559 he had mentioned

Pacioli, and doubtlessly relied on Tartaglia, when he enumerated the possible outcomes of throws of up to four dice and stated the numbers (252 and 462) for five and six dice. He also considered arrangements with repetition by studying combination locks constructed from six rotating rings arranged side by side, each with six letters on it: only one of the $6^6 = 46\,656$ arrangements admitted the opening of the lock.

Christoph Clavius

Christoph Clavius was the chief Jesuit mathematician charged by Pope Gregory XIII with the reform of the Christian calendar. From 1570 he published a long series of mathematical textbooks and commentaries on ancient and medieval writings for mathematical education in the Jesuit colleges.

Christoph Clavius (1538–1612).

His first monograph was *In Sphaeram Ioannis de Sacro Bosco Commentarius* (Commentary on the Sphere of John of Holywood) (see [11]). While explaining the Aristotelian theory of the number and order of the elements, he seized the opportunity to insert a 'most-wonderful digression on the combinations of things'. This digression later influenced an impressive list of authors: Georg Henisch (1608), Paul Guldin (1622), Marin Mersenne (1625), Pierre Hérigone (1634), Daniel Schwenter (1636), Kaspar Schott (1658), Sebastián Izquerdo (1659), Gottfried Wilhelm Leibniz (1666), Athanasius Kircher (1669), Johannes Caramuel (1670), and Kaspar Knittel (1682) (see [19, pp. 6–10]).

Clavius must have been aware of earlier writers, because he said that only a few of them had explained such rules; maybe he was alluding to Cardano, because he had used some of Cardano's examples (planets and letters).

A diagram from Clavius's combinatorial works.

Clavius dealt with just three questions, which he illustrated with several examples:

- He expressed in words the formula $\frac{1}{2}n(n-1)$ for the number $C(n, 2)$ of 2-subsets of an n-set – that is, of combinations in the strict sense of the word.
- He stated that there are $2^n - 1 - n$ 'conjunctions altogether' of n-sets – that is, he understood 'conjunction' or 'combination' in the strict sense of the word: the one 0-subset and the n 1-subsets are not conjunctions, and thus the number $n + 1$ is

to be subtracted from 2^n. Clavius's examples include the four elements, the five predicables (general notions), the seven planets, and the twenty-three letters of the alphabet (with $2^{23} - 1 - 23 = 8\,388\,584$ conjunctions).

- He mentioned that even more expressions (*dictiones*) would appear if the elements in every conjunction could be permuted; this is reminiscent of Cardano's *Practica Arithmeticae*. Thus, he envisioned the notion of arrangements without discussing the general case. He considered only $P(n,n)$, the permutations without repetition, and calculated them by means of the recursion formula

$$(n-1)! \times n = n! \quad \text{or} \quad P(n-1,n-1) \times n = P(n,n);$$

for example, $23! = 25\,852\,016\,738\,884\,976\,640\,000$.

Letters, planets, elements, and seating arrangements of companies at table remained his favourite examples when he illustrated combinatorial operations.

Marin Mersenne

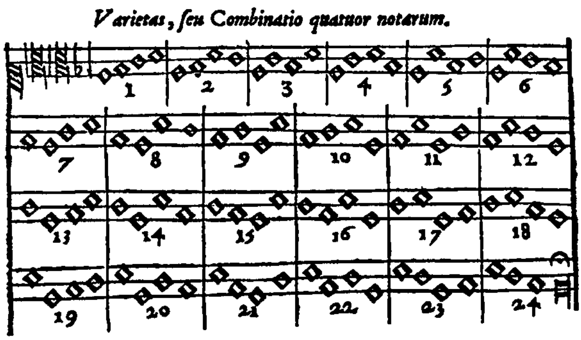
The French Minimite friar Marin Mersenne (1588–1648) was deeply influenced by Llullism and was by far the most important Renaissance author in the history of combinatorics before Gottfried Wilhelm Leibniz (1646–1716). His religious interests were inseparably connected with his interest in the Llullistic combinatorial art, believing that God was the first to practise this fundamental universal art when he created the world, and that God was the first combinatorialist when he combined the single parts of the universe: thus, mankind must imitate God in order to be creative. This especially applied to music: Llullism aimed at musical education, Christianization, God's glorification, creativity, and optimization.

Mersenne's combinatorial studies are set out in six publications, whose titles show that he was addressing mathematicians as well as theologians. In 1623, he published his huge encyclopedia *Quaestiones Celeberrimae in Genesim* (Most Famous Questions Related to Genesis), whose title ends 'A Work Useful for Theologians, Philosophers, Physicians, Legal Advisors, Mathematicians, Musicians, But Especially for Those who are Dealing with Optical Reflections' (see [21]); he used similar formulations again and again.

Two years later there appeared *La Vérité des Sciences contre les Sceptiques ou Pyrrhoniens* (The Truth of the Sciences Against the Sceptics or Pyrrhonians) (see [22]); its aim was to demonstrate that mathematics is most useful for the understanding of the Holy Scriptures.

In both of these monographs Mersenne explained the rules for calculating the numbers of permutations without repetition ($P(n,n) = n!$) and of combinations of n elements ($2^n - 1 - n$), completely relying on Clavius, but his rule for reckoning the number of permutations with repetition was still false. Compared with Clavius, the only novelty was a table for the values 1! to 50! and an enumeration of the 120 songs consisting of the five notes ut, re, mi, fa, sol.

His *Cogita Physico-mathematica* (Physico-mathematical Considerations) of 1644 and his *Novae Observationes Physico-mathematicae* (New Physico-mathematical Observations) of 1647 (see [25] and [26]) contain some additional remarks with regard to the two theoretical books on music, which include Mersenne's most important contributions to combinatorics – his French *Harmonie Universelle Contenant la Théorie et la Pratique de la Musique* (Universal Harmony Containing the Theory and Practice of Music) of 1636 [24] and his Latin *Harmonicorum Libri, in Quibus Agitur de Sonorum Natura, Causis, et Effectibus: De Consonantiis, Dissonantiis, Rationibus, Generibus, Modis, Cantibus, Compositione, Orbisque Totius Harmonicis Instrumentis* (Books about Harmony Dealing with Nature, Causes and Effects of the Notes, With the Consonances, Dissonances, Proportions, Keys, Modes, Songs, Composition, and Harmonical Instruments of the Whole World) of 1635/6 (see [23] and [24]).

On the basis of extensive tables that illustrated all of the above rules and theorems by means of musical notes, he offered a detailed discussion of five of the six basic combinatorial operations: permutations with or without repetition, arrangements with or without repetition, and combinations without repetition. $P(n, n)$ was calculated by means of the above recursion rule.

Combinations of four notes, from Mersenne's *Harmonicorum Libri*.

Mersenne continued the table of such permutations up to 64!, a ninety-digit number and the largest factorial ever calculated up to then, although his table contains a number of errors. He listed the 720 different songs with six notes, writing them out in musical notation, and systematically studied all the different types of repetition regarding the number 9; this was presumably reminiscent of Llull, who had selected nine fundamental notions for his theory of language. In modern terms, he looked for the multinomial numbers $C(n; n_1, n_2, \ldots, n_p)$, where

$$n_1 + n_2 + \cdots + n_p = n \quad \text{with each } n_i \geq 1.$$

Mersenne cited the short treatise *Ars Combinandi* (The Art of Combining) by Jean Matan, who had done the same for $n = 5$.

Arrangements were dealt with as generalizations of permutations, as an ordered selection of p of the n elements, where the formula is

$$P(n, p) = n \times (n - 1) \times \cdots \times (n - p + 1).$$

Mersenne took $n = 22$, and specifically calculated the sum of the different values

$$P(22, 1),\ P(22, 2),\ \ldots\ ,\ P(22, 22).$$

He then did the same for the n^p arrangements with repetition, and calculated $n^p - P(n, p)$, the number of arrangements in which at least one element is repeated.

In order to get the number $C(n, p)$ of unordered selections of p-subsets of an n-set, he used the rule

$$C(n, p) = \frac{P(n, p)}{P(p, p)}.$$

Instead of discussing combinations with repetition in general, he considered the more difficult problem of finding combinations of n symbols that represent a special type of repetition, always relying mainly on his musical examples. As in the case of permutations with repetition, he represented the types by number-theoretical partitions of n. He did not explain his method, which can, however, be reconstructed from the special cases he had discussed earlier.

Let p be the number of selected notes and n be the number of given notes. The partition of p is

$$p = 1r_1 + 2r_2 + \cdots + pr_p, \quad \text{where } 0 \leq r_i \leq p \text{ and } i = 1, 2, \ldots, p$$

– that is, r_i pairwise-distinct notes occurring i times. Altogether, there are $r_1 + r_2 + \cdots + r_m = r$ distinct notes, for $1 \leq m \leq p$.

We thus have $C(n, r)$ combinations. We get a new unordered selection of each given type whenever a note of a certain frequency replaces a note of another frequency – for example, if two notes that occur once or twice are exchanged. We do not get a new combination if two notes of the same frequency are exchanged, because we are dealing with unordered selections. The same frequencies function as repetition in a permutation with repetitions of a certain type. We therefore have to multiply $C(n, r)$ by $r! / r_1!\, r_2! \cdots r_m!$; for example, if nine elements are selected from twenty-two, the partition 2, 2, 1, 1, 1, 1, 1 leads to $C(22, 7) \times 7! / 5!\, 2! = 3\,581\,424$.

In a similar way, Mersenne calculated the number of ordered selections. Here the order of the p selected notes matters, but now the type of repetition of p has to be interpreted as a permutation with repetition. As a consequence, the number of unordered selections has still to be multiplied by $p! / \big((1!)^{r_1} \times (2!)^{r_2} \times \cdots \times (m!)^{r_m}\big)$.

Athanasius Kircher

The most famous Jesuit Llullist and universal scientist of the 17th century was Athanasius Kircher, the German professor of mathematics at the Roman College in Rome, who was called a 'universal genius' and a 'master of a hundred arts'. The 'universal science' was an imitation of God's creative force, which had arranged the world according to measure, number, and weight (Wisdom XI, 20).

Athanasius Kircher (1602–80).

Kircher's universal science was characterized by two main aspects – his analogical thinking and the combinatorial art (see [20] and [32]). There was thus a strong religious context for Kircher's combinatorial approach to mathematizing different sciences, such as music theory, linguistics, and the philosophy of science. In 1663 his conception of a 'new and universal language' (*Polygraphia Nova et Universalis*) was based on combinations of notions in different languages between which he presupposed a one-to-one correspondence. His main contribution to combinatorics is contained in the eighth book of his *Musurgia Universalis sive Ars Magna Consoni et Dissoni in X Libros Digesta qua Universa Sonorum Doctrina, et Philosophia, Musicaeque tam Theoreticae, quam Practicae Scientia, Summa Varietate Traditur* (Universal Musical Art or Great Art of Consonance and Dissonance, Subdivided into Ten Books, by which the Whole Doctrine and Philosophy of Notes and the Science of Theoretical and Practical Music are Treated with the Greatest Versatility) (see [15]). He equated 'to compose' with 'to combine', thereby elaborating a mechanical method of composing that was later denounced as 'sounding algebra'.

Although he did not reveal his source, Kircher completely depended on Mersenne's two great monographs on music theory. He did not discuss all of Mersenne's combinatorial problems, yet he still exhibited a remarkable knowledge of combinatorics. Kircher wanted to demonstrate that the force of numbers and combinations, the huge variety of possible permutations and ordered and unordered selections of relatively few elements, constitutes the universal harmony of God's creation. He even used Mersenne's number examples, with nine and twenty-two elements, and most of his examples were taken from music theory.

Kircher repeated Mersenne's rules for $P(n, n)$, $P(n, k)$, and arrangements with repetition, but calculated the factorials only up to 24!. As to permutations with repetition, Mersenne had explained that there is just one permutation of (say) nine equal notes. Kircher interpreted the notion of 'mutatio' (permutation) in the strict sense of the word, and asserted that there can be no permutation at all of n equal elements. He did not notice that this assertion contradicted the division rule for permutations with repetition, taken over from Mersenne.

Whereas Mersenne had listed all thirty possible types of repetition of nine notes, Kircher considered only nineteen of them, without referring to the number-theoretical context – that is, partitions of 9. The type of repetition $(n - k, k)$ of n elements leads to

$$\frac{n!}{(n-k)!\,k!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = C(n,k);$$

in other words, Kircher calculated $C(n, k)$ by means of the formula for permutations with repetition. But he left aside the most difficult problem of calculating the combinations or arrangements of certain types of repetition. He applied the principle of multiplication of choices to a sequence of bars consisting of different numbers of notes and admitting different numbers of permutations:

If there are k successive choices to be made, and if the ith choice can be made in n_i ways, for $1 \leq i \leq k$, then the total number of ways of making these choices is the product $n_1 \times n_2 \times \cdots \times n_k$.

In 1669 Kircher revisited his mathematical studies related to combinatorics in his *Ars Magna Sciendi sive Combinatoria qua ad Omnium Artium Scientiarumque Cognitionem Brevi Adquirendam Amplissima Porta Recluditur* (The Great Art of Knowledge or the Combinatorial Art, Through which the Broadest Door is Opened for Quickly Acquiring Knowledge in All Arts and Sciences) (see [16]). He mainly repeated Clavius's combinatorial results, adding only the calculation of all factorials up to 50! and special rules for permutations with repetition where only one element is repeated twice, three, four times, etc. This time his illustrations of the combinatorial rules concerned letters and Llullistic notions.

Kircher's pupil and collaborator Kaspar Schott (1608–66) was interested in what Kircher had called the 'force of number and combination'. In Volume 3 of his four-volume *Magia Universalis* (Universal Magic) (1657–59) (see [33]), he referred to Clavius, Kircher, and Tacquet, other members of his order. The 'Arithmetical Magic' (8th book) dealt with permutations and selections, like Schott's three predecessors, but now included arrangements of arrangements. Schott arranged two sets of pairwise distinct elements, thus getting

$$2\,P(n_1, k_1) \times P(n_2, k_2);$$

the factor 2 arises since the two sets can be interchanged.

The Llullists: Izquierdo, Caramuel, and Knittel

In 1659, around the same time as Schott published his four-volume *Magia Universalis*, the Spanish Jesuit and Llullist Sebastián Izquierdo (1601–81) published his *Pharus Scientiarum* (Lighthouse of Sciences) (see [10] and [14]), an encyclopedia of the Llullist conception of science. Its 29th disputation was exclusively dedicated to combinatorics, consisting of 159 paragraphs and including twenty-two tables that differ almost completely from all known tables of his predecessors.

Izquierdo cited only Clavius, but he might also have read Cardano's *Opus Novum de Proportionibus*, because the forms of their arithmetical triangle are identical. But apart from that, Izquierdo wrote a very clear, systematic, and original general treatise on combinatorics, the longest known before those of Leibniz and Jacob Bernoulli. There are still no mathematical demonstrations of the theorems and rules, which are simply illustrated by twenty-two tables and many examples.

First, Izquierdo distinguished between three main groups of combinations: those that differ only by substance (combinations without repetition), by position (permutations without repetition), and by repetition. There is no modern equivalent of the third group, and indeed no successor of Izquierdo (besides Caramuel) took up this classification; this may be because the third type is not based on a consistent notion: if there are n elements $a, b, c, \ldots$, then the total number of combinations differing only by repetition is n^2:

$$a, aa, aaa, \ldots, aa\cdots a \text{ (}n\text{ times)}, \; b, bb, bbb, \ldots, bb\cdots b \text{ (}n\text{ times)}, \; \text{etc.}$$

Each combination contains only one continually repeated element, while different combinations can contain different elements, contradicting the definition. Izquierdo would have had to confine himself to one continually repeated element.

The three main groups lead to four further groups of combinations: those that differ by substance and position (arrangements without repetition), by substance and repetition (combinations with repetition), by position and repetition (permutations with repetition), and by substance, position, and repetition (arrangements with repetition).

To a large extent, Izquierdo presented the combinatorial knowledge we already know from Mersenne. Yet, there are some interesting differences between these two authors. Izquierdo maintained that no rule can be found in other authors for calculating the number of combinations of n different things taken r at a time; possibly, he did not know such predecessors. He explained the additive law of formation of the arithmetical triangle. Instead of combinations of certain types of repetition, he considered combinations with repetition in general, using the arithmetical triangle and the relation

$$C(n+k-1,k)+C(n+k-1,k+1)=C(n+k,k+1).$$

He did not present an independent rule to calculate such combinations directly. He found the number $A(n)$ of all arrangements without repetition of n elements taken r at a time (for $r=1,2,\ldots,n$), according to the recursion

$$A(n)=nA(n-1)+n.$$

Only a few years after Izquierdo, the Spanish Cistercian Juan Caramuel de Lobkowitz (1606–82) published several large volumes partly dealing with combinatorics, metrics (1663) [3], and rhythmics (1665) [4], and especially the two-volume *Mathesis Biceps Vetus et Nova* (Old and New Two-Headed Mathematics) (1670) (see [5]).

A long section in the second volume (Syntagma 6) deals with combinatorics, games of dice, and lotteries. At the beginning, Caramuel referred to Izquierdo when he defined the notion of a combination. The following section on combinatorics is nothing but a complete repetition of Izquierdo's explanations, including his terminology and tables. Caramuel cited Cardano's *Opus Novum de Proportionibus*, where Cardano used the multiplicative law of formation of combinations without repetition. There can be no doubt that Caramuel took it from Cardano, although he maintained that it could not be found in any other author.

Caramuel possibly also knew Cardano's posthumously published *Liber de Ludo Aleae* (1663). In any case, there are some similarities between Cardano's book and Caramuel's section on the game of dice (*Kybeia*). Caramuel correctly listed the possible outcomes of throwing two dice and briefly discussed the game with two astragals. The third section inquired into combinatorial questions in connection with lotteries.

Llullism declined in the second half of the 17th century. Scholars like Kircher or Caramuel were severely criticized by representatives of the new science, such as Leibniz, Nicolaus (I) Bernoulli, and de Montmort. Yet in 1682 the German Jesuit Kaspar Knittel (1644–1702) published his *Via Regia ad Omnes Scientias et Artes* (Royal Road to All Sciences and Arts) [18], whose combinatorial explanations were based mainly on those of other Jesuits, such as Clavius, Guldin, Kircher, and Schott.

Title page of Caramuel's *Mathesis Biceps*.

French and Belgian contributions

Little is known about the life of the French mathematician Pierre Hérigone (d. *c*.1643); maybe his name was a pseudonym for Clément Cyriaque de Mangin. He taught mathematics in Paris, where he published a six-volume Latin–French mathematical encyclopedia from 1634 to 1642, which was reprinted in 1643/4 as 'Mathematical Textbook Demonstrated by a Short and Clear Method by Means of Real and Universal Signs Which are Easily Understandable Without the Use of an Arbitrary Idiom' (see [13]). Doubtless, he based his combinatorial explanations particularly on Clavius, adding some details, the multiplicative calculation of $C(n, k)$, the arithmetical triangle as a tool in the theory of binomial expansions, and the solution of the question

How many combinations without repetitions of n things taken k at a time contain a certain element?

His correct answer was $C(n-1, k-1)$.

In his *Theoria et Praxis Arithmeticae* (Theory and Practice of Arithmetic) (1656/65), the Belgian Jesuit André Tacquet (1612–60) relied on Hérigone and Kircher, explicitly mentioning his sources (see [37]), and repeating Hérigone's question together with its solution, and Kircher's rule for calculating the number of permutations with repetition, and added three observations:

- $C(n, k) = C(n, n-k)$ (Cardano and Mersenne had already formulated this relation);
- the numbers of combinations increase as k approaches $\frac{1}{2}n$;
- if n is even, then $C(n, \frac{1}{2}n)$ is the greatest number; if n is odd, the greatest value occurs twice, for $k = \frac{1}{2}(n \pm 1)$.

Using a lengthy calculation, Tacquet illustrated the gigantic size of $P(n, n)$ for $n = 24$: even if each writer filled forty pages with forty permutations per day, a billion writers could not write down all 24! permutations of 24 letters of the alphabet in a billion years.

Tacquet's *Theoria et Praxis Arithmeticae* was used by Jean Prestet (1652–90), a pupil of Nicolas Malebranche (1638–1715). In 1675 Prestet entered the order of the French 'Oratorians of our Lord Jesus Christ', and in the same year published his *Élémens des Mathématiques ou Principes Généraux de Toutes les Sciences, Qui ont les Grandeurs pour Objets* (Elements of Mathematics or General Principles of All the Sciences whose Objects are Quantities) (1689/95) [30]. Prestet's monograph was an algebra textbook that included some combinatorics. He criticized Kircher's rule for the number of permutations with repetition of just one repeated element and generalized the rule for arbitrarily many repeated elements. As we have seen, Kircher elaborated his relevant table only for one repeated element, but he considered this table as the paradigm for more general types of repetition.

Prestet also discussed arrangements with repetition. He chose as an example the number of numbers that can be formed from $0, 1, 2, \ldots, 9$, and calculated the number of words consisting of at most 24 letters of the alphabet.

The French–Belgian contribution shows that combinatorial studies became an integral part of arithmetical and algebraic textbooks in the 17th century.

Dutch and English contributions

The Dutch mathematician Frans van Schooten the Younger (1615–60) is well known as the editor of François Viète's mathematical works (1646) and translator into Latin of, and commentator on, René Descartes's *La Géométrie* (1649 and 1659/61). In 1657 he published his five books of *Exercitationes Mathematicae* (Mathematical Exercises) [39]: and Gottfried Wilhelm Leibniz (1646–1716), John Wallis (1616–1703), John Kersey (1616–77), Thomas Strode (fl. 1642–88), and Jacob Bernoulli (1654–1705) used them when they wrote their own contributions to combinatorics.

Van Schooten's fifth book is a collection of various problems. To find all selections (or combinations) of a set of n elements is, he said, very similar to the problem of finding all divisors, or aliquot parts (with divisors smaller than m), of a natural number

$$m = p_1 \times p_2 \times \cdots \times p_n,$$

the product of n distinct prime numbers. There are $2^n - 1$ selections or aliquot parts – he ignored the possibility of selecting no element – and 2^n divisors. He also considered special examples of combinations with repetition, but without formulating a general rule.

The same applies to the inverted problem, where he looked for suitable products with a given number of aliquot parts. If k aliquot parts are required, then a^k is a suitable product. Other products come into question, and were mentioned by van Schooten, but he confined himself to certain special examples.

Van Schooten also systematically studied Bachet's weighing problem, a famous problem in additive number theory or the theory of partitions: each natural number can be represented as a sum of distinct powers of 2 if only positive terms are allowed, or as a sum of distinct powers of 3 if negative terms are also allowed.

Towards 1671 John Wallis, Savilian Professor of Geometry at the University of Oxford, elaborated his *Discours of Combinations, Alternations, and Aliquot Parts*, which appeared as an appendix to his 1685 *Treatise of Algebra* (see [34] and [40]), and a Latin version was published in his *Mathematical Works* (1693). Wallis developed van Schooten's ideas and put them into a broader combinatorial context.

He began with an explanation of the construction of the arithmetical triangle, explicitly including the case that with no element the selection can be made in exactly one way. He expressed in words the general rules for the numbers of permutations of n elements with and without repetition, and illustrated them by the ringing of twenty-four bells, and by permuting the letters of words or the words of verses. He mainly relied on Gerhard Johann Voss (1577–1649) and his posthumously published book *On the Four Popular Arts, on Philology, and on the Mathematical Sciences* (1650). In words, Wallis rather clumsily expressed the rule for finding the number of aliquot parts of a product

$$p_1^{a_1} \times p_2^{a_2} \times \cdots \times p_n^{a_n}$$

of prime numbers p_i: this is $\big((a_1 + 1) \times (a_2 + 1) \times \cdots \times (a_n + 1)\big) - 1$, 'which theorem contains the main substance of the *Doctrine of Aliquot Parts*'. It was, however, not new when Wallis published it in 1685: John Kersey had already published it in 1674 in his *Elements of that Mathematical Art Commonly Called Algebra*. Around 1676, Leibniz deduced it without publishing it when he studied van Schooten's *Exercitationes*, rightly saying that van Schooten did not know such a beautiful, short, and general theorem.

Wallis did not mention Thomas Strode's interesting *Short Treatise of the Combinations, Elections, Permutations, and Compositions of Quantities* [36], which appeared in 1678 as the first published monograph on combinations in England. Strode referred to Tacquet's *Theory and Practice of Arithmetic* (1656), to Pascal's *Traité du Triangle Arithmétique* (Treatise on the Arithmetical Triangle) (1665), to van Schooten, and to Prestet's *Elements of Mathematics* (1675), which he falsely attributed to Malebranche. Strode taught the multiplicative calculation of

$$C(n, k) = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

and, like Wallis, calculated the factorials up to 24! when studying permutations without repetition. He then deduced the number of arrangements by means of the formula

$$P(n,k) = C(n,k) \times P(k,k).$$

The game of dice enabled him to study the distinct types of combination with repetition when six elements (the outcomes of throwing six dice) are combined, finding that there are eleven partitions of the number 6. He calculated the possible permutations of such types according to the division rule for permutations with repetition. Interestingly, he also considered other regular polyhedral dice. In his opinion, arrangements with repetition selected from n elements (he called them 'compositions') were the easiest case: their number is n^k.

References

1. J. Borrel (Buteo), *Logistica, Quae et Arithmetica Dicitur in Libros Quinque Digesta*, Lyon (1559).
2. W. Buckley, *Arithmetica Memorativa, sive Compendiaria Arithmeticae Tractatio*, London (1567).
3. J. de Caramuel, *Metametrica*, Rome (1663).
4. J. de Caramuel, *Rhythmica*, San Angelo dei Lombardi (1665).
5. J. de Caramuel, *Mathesis Biceps Vetus et Nova*, Campagna (1670).
6. G. Cardano, *Practica Arithmeticae Generalis*, Milan (1539) = *Opera* IV, 13–216.
7. G. Cardano, *De Subtilitate*, Nuremberg (1550) = *Opera* III, 352–672.
8. G. Cardano, *Opus Novum de Proportionibus*, Basel (1570) = *Opera* IV, 463–601.
9. G. Cardano, *Liber de Ludo Aleae*, in *Opera* I, Lyon (1663), 262–76.
10. R. Ceñal, *La Combinatoria de Sebastián Izquierdo*, Pharus Scientiarum (1659), Disp. XXIX: *De Combinatione, Texto Latino y Traducción Española con una Introdución: La 'Disputatio de combinatione' de Izquierdo en la Historia de la Aritmética Combinatoria, desde Clavius a Bernoulli*, Artes Gráficas Soler, S.A., Madrid (1974).
11. C. Clavius, *In Sphaeram Ioannis de Sacro Bosco Commentarius*, Rome (1570): last (seventh) version in *Opera Mathematica*, Vol. 3, 1st part, Anton Hierat, Mainz (1611) (reprinted by Olms–Weidmann, 1999).
12. R. Haller, Combinatio: nicht Buckley, sondern Cardano – und die $(2^n - 1)$-Regel, *Praxis der Mathematik* 41 (1999), 59–61.
13. P. Hérigone, *Cursus Mathematicus, Nova, Brevi, et Clara Methodo Demonstratus, per Notas Reales et Universales, Citra Usum Cuiuscunque Idiomatis, Intellectu Faciles*, Simeon Piget, Paris (1634–42) (2nd edn, 1643/4).
14. S. Izquierdo, *Pharus Scientiarum, ubi Quidquid ad Cognitionem Humanam Humanitus Acquisibilem Pertinet, Ubertim Iuxta, Atque Succincte Pertractatur*, Claude Bourgeat and Michel Lietard, Lyon (1659).
15. A. Kircher, *Musurgia Universalis*, Haeredes Francisci Corbelletti, Rome (1650).
16. A. Kircher, *Ars Magna Sciendi sive Combinatoria*, Johannes Janssonius a Waesberge, Amsterdam (1669).

17. P. Klopsch (ed.), *Pseudo-Ovidius De Vetula*, Untersuchungen und Text, E. J. Brill (1967).
18. K. Knittel, *Via Regia ad Omnes Scientias et Artes, Hoc est: Ars Universalis Scientiarum Omnium Artiumque Arcana Facilius Penetrandi*, Prague (1682).
19. E. Knobloch, *Die Mathematischen Studien von G. W. Leibniz zur Kombinatorik: Auf Grund fast Ausschließlich Handschriftlicher Aufzeichnungen Dargelegt und Kommentiert*, Steiner Verlag (1973).
20. T. Leinkauf, *Mundus Combinatus, Studien zur Struktur der Barocken Universalwissenschaft am Beispiel Athanasius Kirchers SJ* (*1602–1680*), Akademie Verlag (1993).
21. M. Mersenne, *Quaestiones Celeberrimae in Genesim*, Sébastian Cramoisy, Paris (1623).
22. M. Mersenne, *La Vérité des Sciences*, Paris (1625) (reprinted by Frommann-Holzboog, 1969).
23. M. Mersenne, *Harmonicorum Libri*, Guillaume Baudry, Paris (1635).
24. M. Mersenne, *Harmonie Universelle*, Paris (1636) (reprinted in 1965).
25. M. Mersenne, *Cogita Physico-mathematica*, Paris (1644).
26. M. Mersenne, *Novae Observationes Physico-mathematicae*, Paris (1647).
27. O. Ore, *Cardano, the Gambling Scholar*, with a translation from the Latin of Cardano's *Book on Games of Chance*, by S. H. Gould, Princeton University Press (1953) (reprinted by Dover, 1965).
28. L. Pacioli, *Summa de Arithmetica Geometria Proportioni e Proportionalità*, Venice (1494).
29. E. W. Platzek, *Raimund Lull, Sein Leben – seine Werke, Die Grundlage seines Denkens* (*Prinzipienlehre*), 2 vols., Düsseldorf (1962–4).
30. J. Prestet, *Elémens des Mathématiques*, Paris (1675) (2nd edn., 1689; 3rd edn., 1694).
31. D. M. Robathan (ed.), *The Pseudo-Ovidian De Vetula, Text, Introduction, and Notes*, A. M. Hakkert (1968).
32. W. Schmidt-Biggemann, *Topica Universalis, Eine Modellgeschichte Humanistischer und Barocker Wissenschaft*, F. Meiner (1983).
33. K. Schott, *Magia Universalis Naturae et Artis Opus Quadripartitum*, Johann Gottfried Schönwetter, Würzburg (1657–9).
34. C. J. Scriba, *Studien zur Mathematik des John Wallis* (*1616–1703*), *Winkelteilungen, Kombinatonslehre und Zahlentheoretische Probleme*, Im Anhang die Bücher und Handschriften von Wallis, Steiner Verlag (1966).
35. M. Stifel, *Arithmetica Integra*, Johannes Petreius, Nuremberg (1544).
36. T. Strode, *Short Treatise of the Combinations, Elections, Permutations and Composition of Quantities*, London (1678).
37. A. Tacquet, *Theoria et Praxis Arithmeticae*, Antwerp (1656).
38. N. Tartaglia, *General Trattato di Numeri, et Misure*, Venice (1556).
39. F. van Schooten, *Exercitationes Mathematicae*, Johannes Elzevier, Leiden (1657).
40. J. Wallis, *A Discourse of Combinations, Alternations, and Aliquot Parts*, John Playford, Oxford (1685) (Latin version: *De Combinationibus, Alternationibus, et Partibus Aliquotis, Tractatus, Opera Mathematica*, Vol. 2, Sheldonian Theatre, Oxford (1693), 484–529).

DISSERTATIO
De
ARTE COMBI-
NATORIA,

In qua
Ex Arithmeticæ fundamentis *Complicationum* ac *Transpositionum*
Doctrina novis præceptis exstruitur; et usus ambarum per uni-
versum scientiarum orbem ostenditur; nova etiam
Artis Meditandi,
Seu
Logicæ Inventionis semina
sparguntur.

Præfixa est Synopsis totius Tractatus, et additamenti loco
Demonstratio
EXISTENTIÆ DEI,
ad Mathematicam certitudi-
nem exacta

AUTORE
GOTTFREDO GUILIELMO
LEIBNÜZIO Lipsensi,
Phil. Magist. et J. U. Baccal.

LIPSIÆ,
APUD JOH. SIMON. FICKIUM ET JOH.
POLYCARP. SEUBOLDUM
in Platea Nicolæa,
Literis SPÖRELIANIS.
A. M. DC. LXVI.

Title page of Leibniz's *Dissertatio de Arte Combinatoria*.

CHAPTER 6

The origins of modern combinatorics

EBERHARD KNOBLOCH

In 1654 Fermat and Pascal used combinatorial and other means to solve theoretical questions arising from games of chance; indeed, Pascal's treatise on the arithmetical triangle might be called the first modern treatise on combinatorics. Leibniz was also deeply interested in this subject, but nearly all of his contributions to symmetric functions, partitions, and determinants remained unpublished until recently. Frénicle de Bessy's contributions to combinatorics were also published posthumously. Jacob Bernoulli's posthumously published *Ars Conjectandi* presented an exhaustive treatment of early modern combinatorics. Soon after Bernoulli's death, Pierre Rémond de Montmort and Abraham de Moivre mathematically analysed card games and games of dice in terms of derangements. James Stirling's contributions to combinatorics were motivated by algebraic studies.

Pierre de Fermat and Blaise Pascal

Between July and October 1654 Pierre de Fermat (1601–65) and Blaise Pascal (1623–62) exchanged at least nine letters (two of them lost) on theoretical questions arising from games of chance; this correspondence is now often taken as the beginning of the mathematical theory of probability. They mainly discussed two problems: the *problem of points*, and particular problems involving games of dice. The first of these reads as follows:

> Two players are playing a game, and each player needs a given number of points to win. If they abandon the game before completing it, how should the stakes be divided between them?

Pascal's solution was based on the theory of finite differences. Fermat chose a combinatorial approach that enumerated the possible cases of equal probability – that is, he enumerated arrangements with repetition. If there are two players and four plays left to finish the game, there are 2^4 such arrangements. If there are three players and three plays left, there are 3^3 such arrangements. Interestingly, neither Pascal nor Pierre de Carcavi (*c*.1600–84), with whom he corresponded about Fermat's solution, immediately understood the combinatorial solution (see [1]). Fermat's explanation with regard to games of dice again led to well-known arrangements with repetition: there are 6^3 and 6^4 different throws with three and four dice, respectively.

Combinatorics played no role in Pascal's correspondence with the Belgian canon René François de Sluse (1622–85), even though the latter knew that Pascal occupied himself with this subject. Pascal left behind six studies dealing with combinatorial questions that were motivated by van Schooten's *Mathematical Exercises* of 1657 (see Chapter 5 and [13]), but his results concerning permutations, combinations, and arrangements did not surpass those of his many predecessors, who obviously remained unknown to him.

Probably in 1654 Pascal had finished both a Latin version and a partly French, partly Latin version of a collection of articles, which appeared posthumously in 1665 under the title *Traité du Triangle Arithmétique, avec Quelques Autres Traités sur le Même Sujet* (Treatise on the Arithmetical Triangle, with Some Other Treatises on the Same Subject) (see [18] and Chapter 7). One chapter was dedicated to combinations, but the only new result was $C(n, k) = 0$, for $k > n$. The most important feature of Pascal's *Traité* was his mathematical presentation of well-known results, complete with proofs. Thus, the treatise might be considered to mark 'the beginning of modern mathematical combinatorics' (see [3]).

Gottfried Leibniz

When the German polyhistor, philosopher, and mathematician Gottfried Wilhelm Leibniz wrote his *Dissertatio de Arte Combinatoria* (Dissertation on the Combinatorial Art) [16] in 1666 at the age of 20, he still adhered to Llullism (see Chapter 5), and especially to Athanasius Kircher, who had used the expression 'Ars Combinatoria' in his 1663 *Polygraphia Nova et Universalis ex Combinatoria Arte Detecta* (A New and Universal Writing in Many Languages Revealed by the Combinatorial Art).

However, Leibniz's main interest was in the 'art of inventing', rather than in the mathematical discipline of 'combinatorics'. On the title page he claimed to have proved the existence of God with complete mathematical certainty, because he based his arguments on definitions, axioms, and postulates; the idea of a *mathesis universalis* (universal science) had been crucial to him since the very beginning of his scientific career. Soon after this publication he began to criticize Llullism and to develop a much broader notion of the 'combinatorial art' that not only embraced algebra and the theory of numbers as subdisciplines, but also affected all fields of mathematics known at the time.

Gottfried Wilhelm Leibniz (1646–1716).

Leibniz did not know the relevant mathematical literature when he wrote his dissertation. Mathematically his two main sources were Christoph Clavius (see

Chapter 5) and the *Hours of Mathematical and Philosophical Refreshment* (1636) by Daniel Schwenter [20]. From a modern mathematical point of view, his treatise concerned combinations (Problems 1–3) and permutations (Problems 4–12). He himself admitted that he had given proofs for the solutions of only a few problems, but even these few demonstrations were an essential step towards modern combinatorics (see [9]).

In particular, he calculated the numbers of permutations, combinations, and arrangements without repetition, and in the case of combinations used the additive law of formation

$$C(n,k) = C(n-1,k) + C(n-1,k-1).$$

His false rule for permutations with repetition invalidated his attempt to determine the number of arrangements of a certain type of repetitions, because he multiplied the number of possible combinations of this type by the number of permutations with repetition. Yet, his original method of determining the number of these combinations was correct, and was necessarily identical with Mersenne's expression (see Chapter 5):

Let $p = 1r_1 + 2r_2 + \cdots + pr_p$ be a partition of p, the total number of selected elements; then the number of combinations with this type of repetition is

$$C(n,r_1) \times C(n-r_1,r_2) \times \cdots \times C(n-r_1-r_2-\cdots-r_{p-1},r_p)$$
$$= \frac{n(n-1)\cdots(n-r_1-r_2-\cdots-r_p+1)}{r_1!r_2!\cdots r_p!}.$$

Several other problems that Leibniz dealt with are worth mentioning, because he solved them for the first time. One of these is:

In a set, how many combinations of a certain size, or of all possible sizes, contain a given subset of elements?

He called such a set a 'caput'; Pierre Hérigone and André Tacquet had considered only the special case of a single fixed element. By choosing c objects from a set of n objects, including the caput of m elements, he obtained as his solutions to the two problems $2^{n-m} - 1$ and $C(n-m, c-m)$, respectively; for $m \neq 0$, the first solution should read 2^{n-m}. Leibniz's examples show his creation of the terms '0llio' (nullio) and '1nio' (unio) as early as in his *Ars Combinatoria*. He also coined the word 'super0llio' for $C(m,k) = 0$ where $k > m$.

Leibniz also defined partitions as subsets of combinations. There are $\frac{1}{2}n$ (n even) or $\frac{1}{2}(n-1)$ (n odd) bipartitions of a number n when the order of summands is disregarded, and $n-1$ partitions otherwise. Moreover, Leibniz was already experimenting with partitions involving more than two summands, and partitions later became a favourite subject of his mathematical studies.

How many partitions contain a caput? Leibniz distinguished six possible types of caput, according to whether it has one or several elements, whether it has homogeneous elements that can be placed in a given position in the same way as those already placed, or whether it is monadic (having no homogeneous elements). Thus, in the most general case where, out of n elements, the m elements of the caput can be permuted, the result is

$$(n-m)! \times m! \times C(i_1+a_1, i_1) \times C(i_2+a_2, i_2) \times \cdots \times C(i_k+a_k, i_k),$$

where in the jth case there are i_j $(j = 1, 2, \ldots, k)$ homogeneous elements within the caput that are uniform with a_j outside it.

Finally, Leibniz correctly determined the number of cyclic permutations of n elements as $n!/n = (n-1)!$. Later on, such permutations occurred again in his studies of determinants.

Frénicle de Bessy

The French amateur mathematician Bernard Frénicle de Bessy (*c.*1605–75) left behind several writings that were posthumously published by the French Academy of Sciences [4]. Among these were an *Abrégé des Combinaisons* (Synopsis of Combinations) and a treatise on magic squares; the *Abrégé* originated from Mersenne's books on harmony, which he cited. He first repeated Mersenne's rules for the numbers of permutations and arrangements without or with repetition, and for combinations without repetition. He then declared that the truth can be perceived by means of examples or by consideration. He thus applied both methods, while not writing a mathematical treatise in the modern sense.

Frénicle's examples involved card games, games with three dice, and compound interest. In one of these he explicitly calculated $\left(\frac{21}{20}\right)^{32}$, which is a fraction whose numerator and denominator have 43 and 42 digits, respectively. His rule of 'multiple combinations' described the principle of multiplication of choices. Especially interesting were his linguistic examples, resulting in permutations that are subject to certain restrictions: n letters are permuted, but m of them must not be written side by side, or must not be written at the beginning or end, and so on. Frénicle explained his method: the number of permutations that do not satisfy the restriction must be subtracted from $n!$. Other examples involved the construction of secret writing, the chances of bets, and certain games like chess.

Frénicle's treatise on magic squares was much longer, and dealt with their construction when the order is even and their elements can be permuted. In particular, he included a 'general table', featuring all 880 4×4 magic squares.

Symmetric functions, partitions, and determinants

Leibniz filled hundreds of manuscript pages with studies of symmetric functions, partitions, and determinants, without ever publishing anything on their far-reaching results. The sixty most important sketches on the first two topics were later published in [11], while sixty-seven studies involving the theory of determinants appear in [8] and [12].

Leibniz's studies of symmetric functions (or 'forms', as he called them) were directly connected with his attempts to solve the general polynomial equations of fifth and higher degrees algorithmically. By 1676 he had already found his 'polynomial theorem' – a method for calculating the power $(a_1 + a_2 + \cdots + a_p)^n$. In the early summer of 1678 he corresponded intensively with Ehrenfried Walther von Tschirnhaus (1651–1708) about this algebraic problem. Their correspondence makes evident that at this time Leibniz disposed of the use of tables and rules to determine the number of terms of a form, or to multiply forms with one another. Such tables still exist (see [15]).

At the same time, Leibniz often tried to reduce forms to the simplest possible versions, and especially to elementary symmetric functions. He was convinced that this was always possible and could be done in a unique way when only these functions were used. By this means he formulated the fundamental theorem for symmetric functions, which may now be stated as:

Every polynomial function of the polynomial ring $R[x_1, x_2, \ldots, x_n]$ can be written uniquely as a polynomial in the first n elementary symmetric functions.

For special types of forms, he discovered the laws of formation of the reduced equations – for example,

$$\sum a^3 b^2 cd = \delta yx - 4\varepsilon x^2 + 9\theta x - 21\lambda - 3\delta z + 6\varepsilon y,$$

where $x, y, z, \delta, \varepsilon, \theta$, and λ denote the first seven elementary symmetric functions $\sum a, \sum ab, \sum abc, \sum abcd, \sum abcde, \sum abcdef$, and $\sum abcdefg$.

Around 1678–82, Leibniz occupied himself especially with the calculation of power sums – 'combinatorial powers of a polynomial', as he called them – and found the rule for writing an arbitrary power sum explicitly. Leibniz was the true discoverer of the so-called 'Girard's formula', which Edward Waring (1734–93) had published for the first time in 1762, and which represents power sums exclusively by means of elementary symmetric functions.

After 1700 Leibniz corresponded about this subject with the Wolfenbüttel school principal Theobald Overbeck (d.1719). By 1714 Overbeck had written down an interesting result:

The rule for multiplying power sums. The solution concerns the partition of sets with k elements. Let $k = 1r_1 + 2r_2 + \cdots + kr_n$ be a partition of k. Then there are

$$N = \frac{k!}{(1!)^{r_1}(2!)^{r_2}\cdots(k!)^{r_k} \times r_1!\,r_2!\cdots r_k!}$$

ways of distributing k elements so that one element at a time is put r_1 times in a drawer, two elements at a time r_2 times, . . . , and k elements at a time r_k times.

Overbeck also reduced the multiform symmetric functions to uniform symmetric functions (that is, to power sums), and discovered the first six so-called 'formulas of Waring', which Waring published for the first time only in 1762 (see [10]); these examples show the close connection between Leibniz's studies of symmetric functions and that of number-theoretical partitions.

Let p_n be the number of partitions of a natural number n:

$$n = n_1 + n_2 + \cdots + n_m, \text{ where } 1 \le n_1 \le n_2 \le \cdots \le n_m.$$

Let $p[m, n]$ be the number of partitions of n into m summands, and $p(n, h)$ be the number of partitions of n, the smallest summand of which is h. It was Leibniz's main goal to determine the values of p_n and $p[m, n]$ (see Chapter 9), from which he also hoped to deduce the number of symmetric functions of a certain degree. In the last year of Leibniz's life Overbeck seems to have given him voluminous tables of partition numbers, which led him to the discovery of several laws of recursion.

No later than 1677 Leibniz summed up his solution to $p[2, n]$, given in his *Ars Combinatoria*, as $p[2, n] = \lfloor \frac{1}{2}n \rfloor$. Above all, he tried to determine the value of

$p[3,n]$. Around 1673 he found a rule, erroneously called the 'universal solution', which can be transformed into a formula for $p[3,n]$. Two later attempts failed because of false reasoning or calculation, but the methods were correct, and it is therefore possible to bring Leibniz's various efforts to a successful and improved conclusion.

As early as 1673 Leibniz observed that $p[n,n] = p[n-1,n] - 1$. In the years 1712–15 he found by induction (but did not prove) interesting identities such as

$$p[k,n] = p[k,n-k] + p[k-1,n-1]$$

(Euler's rule of recurrence, published in 1751), $p_n = p[n,2n]$, and

$$p_n = 1 + p(n-1,1) + p(n-2,2) + \cdots + p(n - \lfloor \tfrac{1}{2}n \rfloor, \lfloor \tfrac{1}{2}n \rfloor).$$

Before 1700, Leibniz's attempts to divide forms into a certain number of factors led him to the Stirling numbers of the second kind and to the solution of special problems of permutations. James Stirling (1692–1770) republished these numbers in 1730 (see later).

The combinatorial aspects of the theory of determinants were still so dominant at the beginning of the 20th century that Netto [17] dealt with them under the heading of 'Combinatorics' in the *Encyclopedia of Mathematical Sciences*. Leibniz's interest in determinants was directly connected with his paramount interest in the combinatorial art: it provided the rules according to which characters were to be manipulated to create new knowledge. After his differential and integral calculus, his determinantal calculus is a prime example of how he succeeded in pursuing this basic idea in mathematics. For him, the combinatorial art included algebra, not vice versa, as believed by most mathematicians of his time.

Between 1678 and 1713 Leibniz laid the foundation of the theory of determinants in Europe. No one could imagine what extensive studies Leibniz had pursued on the theory of systems of linear equations and elimination theory, with the aid of expressions that we today call determinants (see [14]). He coined the term 'resultans' (resultant), invented the symbol $\overline{1 \cdot 2 \cdot 3 \cdots n}$ for the $n \times n$ determinant $|a_{ij}|$, and introduced well over fifty different subscript notations for the coefficients of algebraic and differential equations. In 1684 he formulated (without proof) several general theorems concerning combinatorial aggregates (resultants):

- If $\mathbf{A}$ is an $n \times n$ matrix and $d(\mathbf{A})$ is its determinant, then $d(\mathbf{A})$ consists of $n!$ products. Their sign rule is based on the concept of transposition, while the modern definition takes the concept of inversion as its basis.
- When forming the determinant, we may interchange the rows and columns of $\mathbf{A}$.
- If the rows (or columns) $1, 2, \ldots, n$ of a matrix $\mathbf{A}$ are written as a sequence $k_1, k_2, \ldots, k_n$, then the determinant is multiplied by the sign of this permutation.

He also anticipated Laplace's theorem on the expansion of determinants.

Leibniz discovered important results in the theory of systems of linear equations and elimination theory, which he expressed in the language of determinants. In 1684 he discovered 'Cramer's rule' for solving an inhomogeneous system of linear equations. His fundamental treatise has been published in [8]; the Swiss mathematician Gabriel Cramer published it only in 1750.

In 1679–81 Leibniz anticipated James Joseph Sylvester's 'dialytic method' by solving the resultant of two polynomials; Sylvester republished this in 1840. Around 1683–84 Leibniz obtained a method of determining the resultant by means of auxiliary polynomials, later published by Leonhard Euler in 1748 and by E. Bézout in 1764. By 1692–93, Leibniz already knew the most important dimensional and homogeneity properties of the resultant.

Jacob Bernoulli

Jacob Bernoulli was the oldest member of the famous Bernoulli family to be associated with mathematics. With his younger brother Johann (1667–1748), he was one of the earliest and most important propagators of Leibniz's calculus. From his mathematical diary, the *Meditationes*, we know that he occupied himself with games of chance from around 1684. By 1690 he had finished his preparatory work for the *Ars Conjectandi* (Art of Conjecturing), which was posthumously published by his nephew Nicolaus in 1713; for the genesis, history of publication, and an annotated reproduction of this treatise, see Volume 3 of his collected works [2].

Jacob Bernoulli (1654–1705).

As we see in Chapter 7, the *Ars Conjectandi* is divided into four parts. The first part is a reprint of Christiaan Huygens' 1657 treatise *De Ratiociniis in Ludo Aleae* (On Calculations in the Game of Dice) with valuable additions by Bernoulli. The second part is exclusively dedicated to combinatorics, and the third part deals with games of chance. The fourth part applies the calculus of probability to questions in the moral and economic sciences, and includes Bernoulli's proof of his *weak law of large numbers*.

In 1657 Huygens discussed the throwing of two and three distinguishable dice, respectively enumerating the ways of throwing totals of 2 up to 12, and 3 up to 18; such calculations had already been undertaken in the Middle Ages (see Chapter 5). Bernoulli explained how these considerations can be systematically extended to four or more dice, so that no case is forgotten when all essentially different cases are written down.

The second part of the *Ars Conjectandi* is a mathematical textbook on combinatorics, in the modern sense of the word. The content is clearly structured, and the theorems are introduced by examples and rigorously proved. Bernoulli referred to van Schooten, Leibniz, Prestet, and Wallis, showing that he was acquainted with much of the mathematical literature of his subject, but not with authors such as Pascal, Mersenne, or Izquierdo.

While permutations and arrangements with and without repetition, combinations without repetition, and figurate numbers form the familiar core of combinatorial knowledge, Bernoulli seems to be the first to have investigated more difficult questions involving combinations, obtaining the independent expression $C(n+k-1, k)$ for combinations with repetition taken k at a time. He also generalized Leibniz's 'caput' theory:

If we form combinations of n elements taken c at a time, and consider a special subset of m elements (where $m < c$), how many combinations contain exactly b of these m elements?

Bernoulli obtained the answer $C(n-m, c-b)$. Such combinations with restricted repetition correspond to the number of divisors of a product of powers of prime numbers.

Bernoulli developed this topic by dealing with related number-theoretic questions of van Schooten and Wallis, such as:

In how many divisors of a given number does a particular prime number occur? How many divisors have the same number of prime factors?

Bernoulli proved the relation $A(n+1) = nA(n) + 1$, which was already known to Izquierdo (where, as in Chapter 5, $A(n)$ represents the number of arrangements without repetition of n elements taken r at a time, for $r = 1, 2, \ldots, n$).

He also calculated power sums of the form $S(n^c) = 1^c + 2^c + \cdots + n^c$; for example,

$$S(n) = 1 + 2 + \cdots + n = \tfrac{1}{2}n(n+1)$$

and

$$S(n^2) = 1^2 + 2^2 + \cdots + n^2 = \tfrac{1}{6}n(n+1)(2n+1).$$

To this end, he introduced numbers A, B, C, D, ..., later called the *Bernoulli numbers*, defined as the coefficients of n in the expressions for $S(n^2), S(n^4), S(n^6)$, etc. The first four Bernoulli numbers are $\mathrm{A} = \frac{1}{6}, \mathrm{B} = -\frac{1}{30}, \mathrm{C} = \frac{1}{42}$, and $\mathrm{D} = -\frac{1}{30}$, and the general formula reads

$$S(n^c) = \frac{n^{c+1}}{c+1} + \tfrac{1}{2}n^c + \tfrac{1}{2}C(c,1)\mathrm{A}n^{c-1} + \tfrac{1}{4}C(c,3)\mathrm{B}n^{c-3} + \tfrac{1}{6}C(n,5)\mathrm{C}n^{c-5} + \cdots .$$

We note that the first eleven Bernoulli numbers had already been studied by Johann Faulhaber much earlier, between 1612 and 1619.

$$
\begin{aligned}
\int n &= \frac{1}{2} nn + \frac{1}{2} n . \\
\int nn &= \frac{1}{3} n^3 + \frac{1}{2} nn + \frac{1}{6} n . \\
\int n^3 &= \frac{1}{4} n^4 + \frac{1}{2} n^3 + \frac{1}{4} nn . \\
\int n^4 &= \frac{1}{5} n^5 + \frac{1}{2} n^4 + \frac{1}{3} n^3 * - \frac{1}{30} n . \\
\int n^5 &= \frac{1}{6} n^6 + \frac{1}{2} n^5 + \frac{5}{12} n^4 * - \frac{1}{12} nn . \\
\int n^6 &= \frac{1}{7} n^7 + \frac{1}{2} n^6 + \frac{1}{2} n^5 * - \frac{1}{6} n^3 * + \frac{1}{42} n . \\
\int n^7 &= \frac{1}{8} n^8 + \frac{1}{2} n^7 + \frac{7}{12} n^6 * - \frac{7}{24} n^4 * + \frac{1}{12} nn . \\
\int n^8 &= \frac{1}{9} n^9 + \frac{1}{2} n^8 + \frac{2}{3} n^7 * - \frac{7}{15} n^5 * + \frac{2}{9} n^3 * - \frac{1}{30} n . \\
\int n^9 &= \frac{1}{10} n^{10} + \frac{1}{2} n^9 + \frac{3}{4} n^8 * - \frac{7}{10} n^6 * + \frac{1}{2} n^4 * - \frac{1}{12} nn . \\
\int n^{10} &= \frac{1}{11} n^{11} + \frac{1}{2} n^{10} + \frac{5}{6} n^9 * - 1\, n^7 * + 1\, n^5 * - \frac{1}{2} n^3 * + \frac{5}{66} n .
\end{aligned}
$$

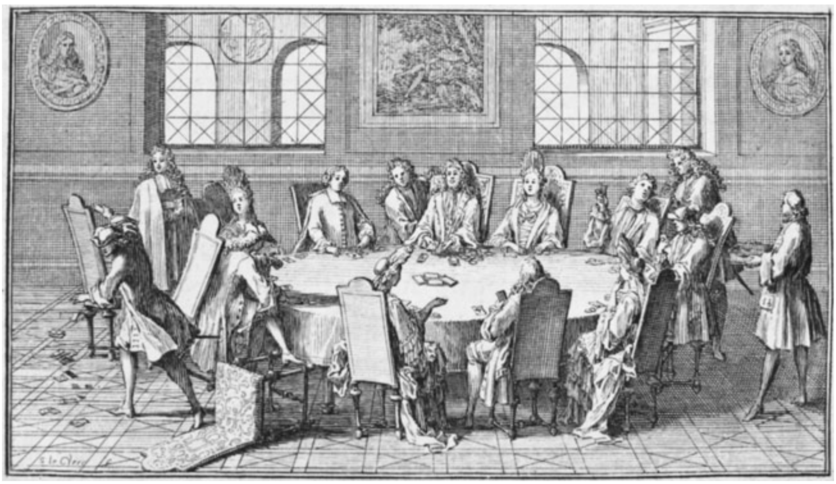
The Bernoulli numbers, from the *Ars Conjectandi*;
the number $-\frac{1}{12}$ at the end of $S(n^9)$ should be $-\frac{3}{20}$.

Pierre de Montmort and Abraham de Moivre

Pierre Rémond de Montmort (1678–1719), the French pupil of Nicolas Malebranche, undertook a systematic exposition of the theory of games of chance (see [23]). His *Essay d'Analyse sur les Jeux de Hazard* (Essay Analysing Games of Chance) was the first mathematical treatise on this subject since Christiaan Huygens' monograph of 1657; it appeared in 1708, five years before the publication of Bernoulli's *Ars Conjectandi*. De Montmort and Nicolaus Bernoulli corresponded about the *Essay*, and their correspondence was included in the 1713 edition, whose first chapter was entitled *Treatise on Combinations*.

In his preliminary remarks de Montmort observed that he had collected the theorems on combinations that were scattered throughout the first edition of the work. Since his main interest concerned card games and games of dice, he did not present a pure account of the subject. His combinatorial explanations were guided by this interest, but he always endeavoured – in contrast to the first edition – to demonstrate the theorems, even though his proofs were sometimes rather cumbersome. In total, there are eighteen theorems and many corollaries.

He first explained the arithmetical triangle, or 'Pascal's triangle', and especially its use for finding the number of combinations. He included the multiplicative law of formation for the numbers $C(n, k)$, and emphasized their connection with the figurate numbers. He also showed how to calculate the sum of the squares, cubes, fourth powers, etc., of the first n natural numbers, and how to find $P(n, k)$, arguing that the coefficients in the expansion of $(a + b)^n$ must equal the number of different ways in which the white and black faces in a gambling game can appear if n counters have been thrown, each counter having one black and one white face. The multinomial theorem is dealt with as a generalization of the binomial theorem.



An illustration from de Montmort's *Essay d'Analyse sur les Jeux de Hazard*.

De Montmort's games of chance included the following:

- *There are thirteen white, black, red, and blue cards. In how many ways can one draw four cards, one of each colour?*
- *Given p dice, each with f faces, how many different throws are there?*
  (The answer f^p concerns arrangements with repetition. This question had been already raised by Tartaglia, and de Montmort showed the connection between its answer and the diagonals of the arithmetical triangle.)
- *In how many ways is the sum of the numbers on the dice equal to a given number n?*
  (If $n - p = s$, then the required number is

$$\frac{p(p+1)\cdots(p+s-1)}{s!} - p\frac{p(p+1)\cdots(p+s-f-1)}{(s-f)!}$$
$$+\frac{p(p-1)}{2!}\cdot\frac{p(p+1)\cdots(p+s-2f-1)}{(s-2f)!} - \cdots,$$

where the series continues as long as all the factors are positive.)

The third of these problems and its solution were additions to the second edition and led to a priority dispute with Abraham de Moivre, who had published the problem in 1711 and its solution in 1730. However, de Montmort already knew it by 1710, as confirmed by a letter he had sent to Johann Bernoulli.

De Montmort's discussion of the card game *Treize* is especially worth mentioning, because it concerns *derangements*, or permutations in which no coincidence occurs (see also Chapter 13): in a permutation of $1, 2, \ldots, n$, a coincidence occurs at the ith place if the ith element is i. In 1708 de Montmort had stated without proof (see [6]) that

$$D_n = (n-1)(D_{n-1} + D_{n-2}), \text{for } n \geq 2,$$

where D_n is the number of permutations of $1, 2, \ldots, n$ in which no coincidence occurs. He concluded that

$$D_n = n!\left(1 - 1/1! + 1/2! - 1/3! + \cdots + (-1)^n/n!\right).$$

In his second edition de Montmort gave two demonstrations which he had received from Nicolaus Bernoulli [6, pp. 301–2]. His first proof used the method of inclusion–exclusion (see Chapter 13); for a detailed discussion see Takács [22], and de Montmort's combinatorial contributions are discussed by Henny [7].

Abraham de Moivre (1667–1754) was one of many Protestants who emigrated from France to England following the revocation of the Edict of Nantes in 1685. In 1711 he published his Latin written treatise *De Mensura Sortis*; expanded versions in English appeared in 1718, 1738, and 1756, under the title *The Doctrine of Chances*. It was his most important work and was the third systematic treatise on probability, after those of Huygens and de Montmort; de Moivre was acquainted with both of them. As for combinatorics, he said in his preface to the first English edition:

One of the principal Methods I have made use of ... has been the Doctrine of Combinations, taken in a sense somewhat more intensive, than as it is commonly understood. The Notion of Combinations being so well fitted to the Calculation of Chance, that it naturally enters the Mind whenever any Attempt is made towards the Solution of any Problem of that kind … The general theorem invented by Sir Isaac Newton, for raising a Binomial to any Power given, facilitates infinitely the Method of Combinations, representing in one view the combination of all the chances, that can happen in any given number of times.

In other words, de Moivre was even more interested than de Montmort in combinatorics applied to games, rather than in the combinatorics itself.

The ninety problems of the second edition (1738) of *The Doctrine of Chances* were preceded by ten cases dealing with, among other things, the probability of throwing one or two 1s in two, three, or four dice throws, or of throwing one 1 and no more in four dice throws. As a consequence, he had to count the chances of such events – that is, to solve combinatorial problems.

Referring to his Latin treatise, but without mentioning de Montmort, he inserted a lemma after his third problem which determined the number of ways of throwing a given number of points with any number of dice. Combinations without repetition, and arrangements with and without repetition, arose in the context of finding the probability of making an unordered or ordered selection of elements from a given set of elements that may be repeated.

Without hinting at de Montmort or Nicolaus Bernoulli, de Moivre discussed the card game *Treize* in greater generality than his predecessors, thereby deriving the above formula for the number D_n of derangements of n objects by the principle of inclusion–exclusion (see Chapter 13 and [5, pp. 95–103]). De Moivre's life and work are discussed in [19].

James Stirling

In 1730 the Scottish mathematician James Stirling (1692–1770) published his *Methodus Differentialis* (Method of Differentials), which deals with the summation and interpolation of infinite series. Under the heading 'On the form and reduction of series', he explained how to express powers of a variable z in terms of multiples of $z-1, z-2, \ldots$, and $z-n+1$; for example,

$$z^3 = z + 3z(z-1) + z(z-1)(z-2).$$

The numerical coefficients were later called *Stirling numbers of the second kind*. They form 'Stirling's triangle', which he himself called his 'first table' (see [21, p. 8]):

$n=$	1	2	3	4	5	6	$\cdots$
$m=1$	1	1	1	1	1	1	$\cdots$
$=2$		1	3	7	15	31	$\cdots$
$=3$			1	6	25	90	$\cdots$
$=4$				1	10	65	$\cdots$
$=5$					1	15	$\cdots$
$=6$						1	$\cdots$

We denote them by $S(n, m)$, the number of partitions of n objects into m classes; for example, $S(4, 2) = 7$, because there are seven partitions of the four elements a, b, c, d into two classes:

$$a|bcd,\ \ b|acd,\ \ c|abd,\ \ d|abc,\ \ ab|cd,\ \ ac|bd,\ \text{and}\ ad|bc.$$

Stirling himself did not speak of any combinatorial context, nor did he mention the recursion law

$$S(n+1, k) = S(n, k-1) + kS(n, k).$$

Leibniz had discovered these numbers earlier, but without identifying them.

There is just as little combinatorial context with regard to Stirling's discovery of his numbers of the first kind (see [21, p. 10]):

$k=$	1	2	3	4	5	$\cdots$
$n=1$	1					
$=2$	1	1				
$=3$	2	3	1			
$=4$	6	11	6	1		
$=5$	24	50	35	10	1	
$=6$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Stirling introduced them in order to express the inverted powers of z, such as $1/z^2$:

$$\frac{1}{z^2} = \frac{1}{z(z+1)} + \frac{1}{z(z+1)(z+2)} + \frac{2}{z(z+1)(z+2)(z+3)} + \frac{6}{z(z+1)(z+2)(z+3)(z+4)} + \cdots.$$

These *Stirling numbers of the first kind*, denoted by $s(n, k)$, are now defined as the coefficients of the polynomal $[x]_n$ of degree n:

$$[x]_n = s(n, 0) + s(n, 1)x + s(n, 2)x^2 + \cdots + s(n, k)x^n,$$

where

$$s(k+1,k) = s(n,k-1) - ns(n,k);$$

thus, they are alternately positive and negative:

$k =$	1	2	3	4	5	$\cdots$
$n = 1$	1					
$= 2$	-1	1				
$= 3$	2	-3	1			
$= 4$	-6	11	-6	1		
$= 5$	24	-50	35	-10	1	
$= 6$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

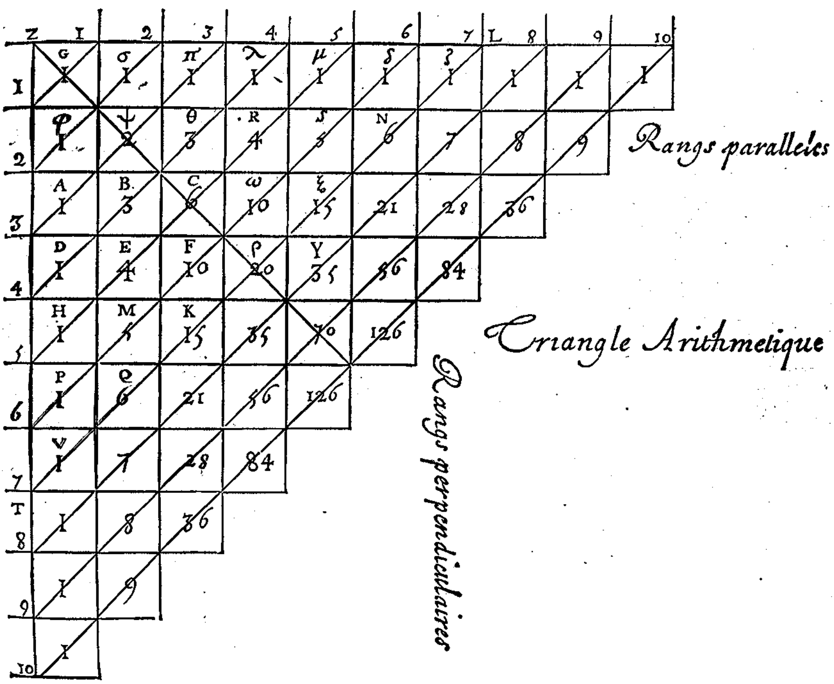
Modern combinatorics came into being during the roughly eighty-year period between 1654 and 1730. Fermat and Pascal's correspondence took place in 1654, and Stirling's *Methodus Differentialis* appeared in 1730, while Leibniz's own life, from 1646 to 1716, spanned much of this developmental period. In general, apart from Jacob Bernoulli, mathematicians of this era did not yet study combinatorics for its own sake, but only did so in order to solve problems regarding games of chance and card games, or algebra and differential calculus. Furthermore, many results were not even published during the lifetime of the authors. This was especially true of Leibniz, who anticipated many mathematical achievements that were then rediscovered many decades after his death.

References

1. P.-J. About and M. Boy, La correspondence de Blaise Pascal et de Pierre de Fermat, *La Géométrie du Hasard ou Le Début du Calcul des Probabilités*, Les Cahiers de Fontenay N. 32 (September 1983).
2. Jacob Bernoulli, *Ars Conjectandi, Opus Posthumum. Accedit Tractatus de Seriebus Infinitis et Epistola Gallice Scripta de Ludo Pilae Reticularis* (ed. N. Bernoulli), Brothers Thurneisen, Basel (1713) (reprinted, together with the relevant sections from J. Bernoulli's mathematical diary *Meditationes*, in Jakob Bernoulli, *Die Werke*, Vol. 3, Naturforschende Gesellschaft in Basel, Birkhäuser (1975), 107–286).
3. N. L. Biggs, The roots of combinatorics, *Historia Math.* 6 (1979), 109–36.
4. Frénicle de Bessy, Abrégé des combinations. Des quarrez magiques, *Divers Ouvrages de Mathématiques et de Physique* (ed. P. de la Hire), Imprimerie Royale, Paris (1693), 45–64, 423–507.

5. A. de Moivre, *The Doctrine of Chances: or, A Method of Calculating the Probabilities of Events in Play*, 2nd edn., H. Woodfall, London (1738) (reprinted by Frank Cass, 1967).
6. P. R. de Montmort, *Essay d'Analyse sur les Jeux de Hazard*, Paris (1708); 2nd edn., J. Quillau, Paris (1713) (reprinted by Chelsea, 1980).
7. J. Henny, Niklaus und Johann Bernoullis Forschungen auf dem Gebiet der Wahrscheinlichkeitsrechnung in ihrem Briefwechsel mit Pierre Rémond de Montmort, *Jakob Bernoulli, Die Werke*, Vol. 3, Naturforschende Gesellschaft in Basel, Birkhäuser (1975), 462–78.
8. E. Knobloch, Die entscheidende Abhandlung von Leibniz zur Theorie linearer Gleichungssysteme, *Studia Leibnitiana* 4 (1972), 163–80.
9. E. Knobloch, *Die Mathematischen Studien von G. W. Leibniz zur Kombinatorik, auf Grund fast Ausschliesslich Handschriftlicher Aufzeichnungen Dargelegt und Kommentiert*, Steiner (1973).
10. E. Knobloch, The mathematical studies of G. W. Leibniz on combinatorics, *Historia Math.* 1 (1974), 409–30.
11. E. Knobloch, *Die Mathematischen Studien von G. W. Leibniz zur Kombinatorik*, Textband, im Anschluss an den gleichnamigen Abhandlungsband zum ersten Mal nach den Originalhandschriften herausgegeben, Steiner (1976).
12. E. Knobloch, *Der Beginn der Determinantentheorie, Leibnizens Nachgelassene Studien zum Determinantenkalkul*, Gerstenberg (1980).
13. E. Knobloch, Sur quelques études mathématiques et physiques de Sluse, *Bull. Soc. Roy. Sci. Liège* 55 (1986), 165–89.
14. E. Knobloch, First European theory of determinants, *Gottfried Wilhelm Leibniz, The Work of the Great Universal Scholar as Philosopher Mathematician Physicist Engineer* (ed. K. Popp and E. Stein), Schlütersche Verlagsanstalt (2000), 55–64.
15. E. Knobloch, The notion of variation in Leibniz, *A Master of Science History: Essays in Honor of Charles Coulston Gillispie* (ed. J. Z. Buchwald), Springer (2012), 235–52.
16. G. W. Leibniz, *Dissertatio de Arte Combinatoria, in qua ex Arithmeticae Fundamentis Complicationum ac Transpositionum Doctrina Novis Praeceptis Exstruitur, et Usus Ambarum per Universum Scientiarum Orbem Ostenditur, Nova etiam Artis Meditandi seu Logicae Inventionis Semina Sparguntur*, Fick & Seubold, Leipzig (1666) (reprinted in *G. W. Leibniz, Sämtliche Schriften und Briefe*, Vol. 1, Akademie der Wissenschaften der DDR, Series VI, 2nd edn., Berlin (1990), 163–230).
17. E. Netto, Kombinatorik, *Encyclopädie der Mathematischen Wissenschaften mit Einschluss ihrer Anwendungen*, Vol. 1, Part 1, Teubner, Leipzig (1898–1904), Chapter A, Arithmetik, Part 2, 29–46.
18. B. Pascal, *Traité du Triangle Arithmétique avec Quelques autres Traitez sur la mesme Matière*, Guillaume Desprez, Paris (1665) (reprinted in *Oeuvres*, Vol. III, L. Brunschvicg and P. Boutroux, Hachette (1908), 433–598).
19. I. Schneider, Der Mathematiker Abraham de Moivre, *Arch. Hist. Exact Sci.* 5 (1968/9), 177–317.

20. D. Schwenter, *Deliciae Physico-Mathematicae oder Mathematische und Philosophische Erquickstunden*, Dümler, Nürnberg (1636).
21. J. Stirling, *Methodus Differentialis: sive Tractatus de Summatione et Interpolatione Serierum Infinitarum*, G. Bowyer, London (1730).
22. L. Takács, On the 'Problem des ménages', *Discrete Math.* 36 (1981), 289–98.
23. I. Todhunter, *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace*, Cambridge (1865) (reprinted by Chelsea, 1965).



Pascal's original arithmetical triangle, as it appears in the frontispiece of his *Traité du Triangle Arithmétique*, written in 1654 and published posthumously in 1665.

CHAPTER 7

The arithmetical triangle

A. W. F. EDWARDS

The arithmetical triangle is the most famous of all number patterns. Apparently a simple listing of the binomial coefficients, it contains the triangular and pyramidal numbers of ancient Greece, the combinatorial numbers that arose in the Hindu studies of arrangements and selections, and (barely concealed) the Fibonacci numbers from medieval Italy. It reveals patterns that delight the eye, raises questions that tax the number-theorists, and amongst the coefficients 'There are so many relations present that when someone finds a new identity, there aren't many people who get excited about it any more, except the discoverer!' [1].

Permutations and combinations

As we have seen in earlier chapters, many combinatorial enumeration problems involve the combinatorial numbers $C(n, r)$, for $n = 1, 2, \ldots$ and $r = 0, 1, \ldots, n$, because these enumerate both the *combinations* of r things selected from n different things, and the *arrangements* or *permutations* of r things of one kind and $n - r$ of another kind. The notation indicates that these numbers are both *combinatorial numbers* and *binomial coefficients*, and a moment's reflection shows the connection between them, for when we expand the binomial expression $(a + b)^n$ the coefficient of $a^r b^{n-r}$ enumerates the number of arrangements of r as and $n - r$ bs.

When the coefficients are arranged in successive rows for each n, the arrangement is known as the *arithmetical triangle*, after the title of Pascal's book *Traité du Triangle Arithmétique* (Treatise on the Arithmetical Triangle). The common arrangement shown below is not the one that Pascal used as the frontispiece to his book, where the coefficients for each n are displayed as diagonals, but it has nevertheless become customary to refer to it as *Pascal's triangle*. As we have seen, there is no implication in this usage that Pascal was the earliest to record it; the eponym arises rather because Pascal was the first author to explore its properties in a systematic manner, identifying the binomial and combinatorial numbers with the *figurate numbers* (natural numbers, triangular numbers, tetrahedral numbers, . . .) of antiquity. For more information about the arithmetical triangle, see [2].

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad .
\end{array}
$$

The usual form of Pascal's triangle.

The combinatorial numbers in India

The connection between the arithmetical triangle and combinatorial problems was first made in India. As we saw in the Introduction and Chapter 1, Piṅgala, a writer on prosody who flourished around 200 BC, gave a rule, by all accounts very cryptically, for finding the number of combinations of n syllables, each of which could be either short or long, when these are taken one at a time, two at a time, three at a time, . . . , all at a time. It seems to have amounted to the observation that the natural numbers give the answers to the first question for successive values of n, the triangular numbers give the answers to the second question, the tetrahedral numbers give the answers to the third question, and so on, through the ascending orders of the figurate numbers. The successive orders of the figurate numbers are given by the rows of Pascal's format for the arithmetical triangle, each row being formed from its predecessor by summation, implying the well-known addition rule for binomial coefficients,

$$C(n+1, r+1) = C(n, r) + C(n, r+1).$$

Piṅgala's rule, known as the *meru prastāra* (the holy mountain), was most succinctly given by his commentator Varāhamihira, who in AD 505 wrote:

It is said that the numbers are obtained by adding each with the one which is past the one in front of it, except the one in the last place.

In the 10th century, Halayudha explained the rule in a way that corresponds to the usual modern form of the arithmetical triangle:

Draw a square. Beginning at half the square, draw two other similar squares below it; below these two, three other squares, and so on. The marking should be started by putting 1 in the first square. Put 1 in each of the two squares of the second line. In the third line put 1 in the two squares at the ends and, in the middle square, the sum of the digits in the two squares lying above it. In the fourth line put 1 in the two squares at the ends. In the middle ones put the sum of the digits in the two squares above each. Proceed in this way. Of these lines, the second gives the combinations with one syllable, the third the combinations with two syllables, etc.

This rule therefore develops the arithmetical triangle, using the addition formula to give the number of permutations of r things of one kind and $n - r$ of another.

The rule for the number of combinations of r things taken from n different things,

$$C(n, r) = \frac{n}{1} \times \frac{n-1}{2} \times \frac{n-2}{3} \times \cdots \times \frac{n-r+1}{r},$$

was given algorithmically by the Jain mathematician Mahāvīra in *Gaṇitasāra-saṅgraha* (Epitome of the Essence of Calculation), written in 850. This rule was repeated in the famous *Līlāvatī* of the Hindu mathematician Bhāskara II in 1150, who took $n = 6$ as an example: six tastes that are to be combined in all possible ways (sweet, pungent, astringent, sour, salty, and bitter). To apply the rule he set down the numbers 1, 2, 3, 4, 5, 6, forwards and backwards in the pattern

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

and by successive application of the rule found the numbers of preparations that combine these six tastes to be 6, 15, 20, 15, 6, and 1, respectively. In so doing he listed the row of the arithmetical triangle for $n = 6$, lacking only the initial 1 corresponding to tastelessness. In another example he generated the row for $n = 8$.

Bhāskara also knew that $C(n, r)$ gives the number of permutations of r things of one kind and $n - r$ of another, using as an example the enumeration of the arrangements of six syllables of which a given number are short and the remainder long, and in this case he noted that we must not forget the arrangement 'all short', making 64 arrangements in all. But he did not give the additive rule of the *meru prastāra*, and thus did not reveal how to generate a row of the arithmetical triangle from its predecessor.

However, another commentator, Bhaṭṭotpala (1068), gave an example involving the combination of sixteen things, explicitly generating the arithmetical triangle, and added that the number of combinations could be found by either rule – that is, by successive additions, or by the above 'multiplicative formula' for $C(n, r)$. Not until 1570 was this connection to be noted in the West. Here is Bhaṭṭotpala's example illustrating the *meru prastāra* rule (this table also appears in Chapter 1 as given by Varāhamihira):

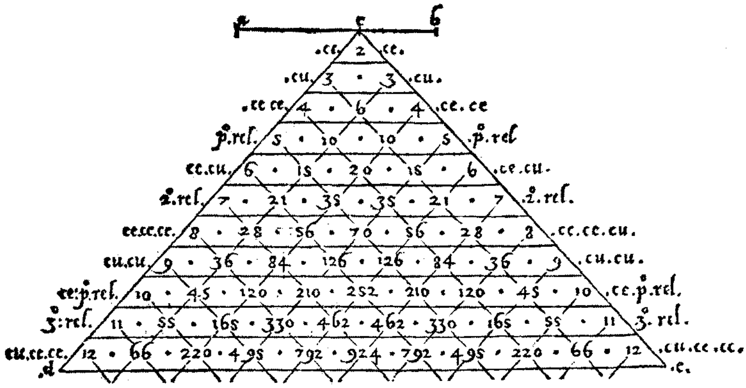
	Taken two at a time	Taken three at a time	Taken four at a time
16			
15	120		
14	105	560	
13	91	455	1820
12	78	364	1365
11	66	286	1001
10	55	220	715
9	45	165	495
8	36	120	330
7	28	84	210
6	21	56	126
5	15	35	70
4	10	20	35
3	6	10	15
2	3	4	5
1	1	1	1

By the start of the second millennium, therefore, we find in India the combinatorial numbers derived for the two simplest problems of permutations and combinations by the two algorithmic routes of multiplication and addition, the latter method generating the arithmetical triangle. In *Līlāvatī* Bhāskara went further, giving the multinomial coefficient for the number of arrangements when there are more than two kinds of things to choose from.

Combinatorial numbers in the West before Pascal

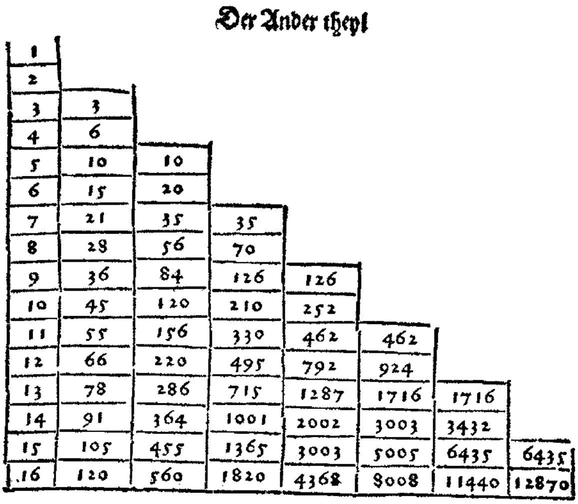
As we saw in Chapter 4, Levi ben Gerson, who lived in France, wrote on permutations and combinations in 1321, and gave in words the multiplicative formula for $C(n, r)$ for the number of combinations of n things taken r at a time, deriving the result directly from the number of arrangements of n things taken r at a time, divided by the number of arrangements of r things. Before then, there had been several examples of Islamic arithmetical triangles (see Chapter 3). But it was not until the 16th century that the arithmetical triangle made its combinatorial debut in the West, in the *General Trattato di Numeri et Misure* (General Treatise on Numbers and Measures) of Niccolò Tartaglia.

Tartaglia sought the number of possible combinations when a number of six-sided dice are thrown. Occasional earlier enumerations had not revealed the essential structure of the solution, but Tartaglia found the connection with the figurate numbers 'on the first day of Lent, 1523, in Verona', as he proudly tells us, 'having thought about the problem all night'. His *General Trattato* was published in 1556, and gives the solution as the first six columns of an arithmetical triangle in Pascal form. Tartaglia probably obtained his result by a clever ordering of the possibilities which facilitated a systematic enumeration, leading to the figurate numbers. He clearly knew the addition rule and the identity between the figurate numbers and the binomial coefficients. He gave the more usual form of the arithmetical triangle for them later in his book; in Italy, 'Pascal's triangle' is sometimes known as 'Tartaglia's triangle'.



Tartaglia's arithmetical triangle.

Meanwhile, in Germany, in 1544, Michael Stifel had presented a form of the arithmetical triangle based on the figurate numbers in connection with the extraction of roots, for which he needed to find the binomial numbers. His method was to write the natural numbers 1, 2, 3, 4, . . . in the first column. Then, in the second column, beginning next to the 3, he wrote the triangular numbers 3, 6, 10, 15, . . . (and he explicitly named them as such). For the third column, next to the 10, he then wrote the tetrahedral numbers 10, 20, 35, 56, . . . – and so on, using the additive procedure for the triangle.



Der Ander theyl

1							
2							
3	3						
4	6						
5	10	10					
6	15	20					
7	21	35	35				
8	28	56	70				
9	36	84	126	126			
10	45	120	210	252			
11	55	156	330	462	462		
12	66	220	495	792	924		
13	78	286	715	1287	1716	1716	
14	91	364	1001	2002	3003	3432	
15	105	455	1365	3003	5005	6435	6435
16	120	560	1820	4368	8008	11440	12870

Stifel's triangle of figurate numbers.

From 1570, when Cardano's *Opus Novum de Proportionibus Numerorum* (New Work on the Proportions of Numbers) appeared, the combinatorial application of the arithmetical triangle entered the mainstream of mathematics. Cardano gave the rule for getting from $C(n, r-1)$ to $C(n, r)$, known to the Hindus, thus generating the multiplicative formula as well as the identification with the figurate numbers and their additive property.

1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1
2	3	4	5	6	7	8	9	10	11	
3	6	10	15	21	28	36	45	55		
4	10	20	35	56	84	120	165			
5	15	35	70	126	210	330				
6	21	56	126	252	462					
7	28	84	210	462						
8	36	120	330							
9	45	165								
10	55									
11										

Cardano's arithmetical triangle.

Tabella pulcherrima & vtilißima Combinationis duodecim Cantilenarum.

I.	II.	III.	IV.	V.	VI.	VII.	VIII.	IX.	X.	XI.	XII.
1	1	1	1	1	1	1	1	1	1	1	1
2	3	4	5	6	7	8	9	10	11	12	13
3	6	10	15	21	28	36	45	55	66	78	91
4	10	20	35	56	84	120	165	220	286	364	455
5	15	35	70	126	210	330	495	715	1001	1365	1820
6	21	56	126	252	462	792	1287	2002	3003	4368	6188
7	28	84	210	462	924	1716	3003	5005	8008	12376	18564
8	36	120	330	792	1716	3432	6435	11440	19448	31824	50388
9	45	165	495	1287	3003	6435	12870	24310	43758	75582	125970
10	55	220	715	2002	5005	11440	24310	48620	92378	167960	293930
11	66	286	1001	3003	8008	19448	43758	92378	184756	352716	646646
12	78	364	1365	4368	12376	31824	75582	167960	352716	705432	1352078
13	91	455	1820	6188	18564	50388	125970	293930	646646	1352078	2704156
14	105	56[illegible]	2380	8568	27132	77520	203490	497420	1144066	2496144	5200300
15	120	680	3060	11628	38760	116280	319770	817190	1961256	4457400	9657700
16	136	816	3876	15504	54264	170544	490314	1307504	326876[illegible]	7726160	17383860
17	153	969	4845	20349	74613	245157	735471	2042975	5311735	13037895	30421755
18	171	1140	5985	26334	100947	346104	1081575	3124550	8436285	21474180	51895935
19	190	1330	7315	33649	134596	480700	1562275	4686825	13123110	34597290	86493225
20	210	1540	8855	42504	177100	657800	2220075	6906900	20030010	54627300	141120525
21	231	1771	10626	53130	230230	888030	3108105	10015005	30045015	84672315	225792840
22	253	2024	12650	[illegible]5780	296010	118404[illegible]	4292145	14307150	44352165	129024480	354817320
23	276	2300	14950	80730	376740	1560780	5852925	20160075	64512290	193536720	548354040
24	300	2600	1755[illegible]	[illegible]8280	475020	203580[illegible]	7888725	28048800	92561040	286097760	834451800
25	325	2925	2047[illegible]	[illegible]8755	593775	2629575	10518300	38567100	131128140	417225900	1251677700

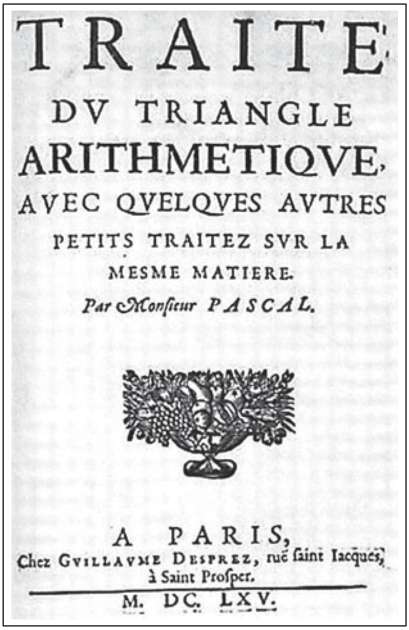
Mersenne's arithmetical table.

By 1636 Father Marin Mersenne had learnt everything that Cardano had written on combinations, and in his *Harmonicorum Libri XII* (Twelve Books of Harmonic Principles) published the largest arithmetical triangle extant: twenty-five rows and twelve columns. Mersenne was familiar with the multiplicative formula and with its combinatorial derivation given by ben Gerson and, as we saw in Chapter 5, applied his combinatorial knowledge to the permutations and combinations of musical notes. The Pascals, father and son, visited him, and it is not surprising that Blaise Pascal's format for the arithmetical triangle is the same as Mersenne's, as Cardano's, and ultimately as Tartaglia's. To Pascal's use of the arithmetical triangle we now turn.

Pascal's *Traité du Triangle Arithmétique*

Pascal's treatise on the arithmetical triangle is recognizably modern. He first established the properties of the binomial coefficients (as the entries of the arithmetical triangle are now universally called) and then, in several appendices, he showed how they could be used to solve a number of mathematical problems. The *Traité* was written in 1654, but was not distributed until 1665, after Pascal's death, when the printed sheets were found amongst his papers; a complete description of Pascal's book is given in [2, Ch. 6 and 7]. The appendix that concerns us is 'Usage du triangle arithmétique pour les combinaisons' (Use of the arithmetical triangle for combinations). It is preceded by an account of the figurate numbers, and is followed by Pascal's famous solution to the *problem of points* – a gambling problem concerning the division of stakes between two players when a game is left unfinished (see Chapter 6), thus introducing the notion of expectation.

Blaise Pascal (1623–1662).


TRAITE
DV TRIANGLE
ARITHMETIQVE,
AVEC QVELQVES AVTRES
PETITS TRAITEZ SVR LA
MESME MATIERE.
Par Monsieur PASCAL.

A PARIS,
Chez GVILLAVME DESPREZ, ruë saint Iacques,
à Saint Prosper.
M. DC. LXV.

Title page of Pascal's *Traité du Triangle Arithmétique*.

Pascal starts:

The word *Combination* has been used in many different senses, so that to avoid ambiguity I am obliged to say what I mean by it

– and he gives its modern meaning. After some further preliminary remarks he presented his important Lemma 4:

the number of combinations of $n + 1$ things taken $r + 1$ at a time is equal to the sum of the number of combinations of n things taken r at a time and the number of combinations of n things taken $r + 1$ at a time

– that is,

$$C(n + 1, r + 1) = C(n, r) + C(n, r + 1).$$

For, said Pascal, using an example,

consider any particular one of the $n + 1$ things: $C(n, r)$ gives the number of combinations that include it whilst $C(n, r + 1)$ gives the number that exclude it, the two numbers together giving the total.

Pascal may well have seen the first part of this reasoning in Pierre Hérigone's *Cours Mathématique* (Course in Mathematics) of 1634.

Having thus established the addition formula by a direct combinatorial argument, Pascal pointed out that the same formula generates the arithmetical triangle, since the initial conditions correspond as well, and therefore that the triangle can be used to solve combinatorial problems.

> *Conclusion*: By the *rapport* [Pascal's word] which exists between the elements of the arithmetical Triangle and combinations, it is easy to see that everything which has been proved for the one applies to the other in like manner, as I shall show in a little treatise I have done on Combinations.

This is presumably his associated Latin treatise *Combinationes*, the first part of which is a Latin version of the above, whilst the remainder interprets some of the 'consequences' of the arithmetical triangle given in the first part of the *Traité du Triangle Arithmétique* in the language of combinations. Although it contains no surprises, it stands as the first systematic application of the arithmetical triangle to combinatorial problems.

Pascal's solution to the problem of points follows as the next section of his *Traité*. Although it involves the arithmetical triangle, its great importance in the history of probability is not due to the solution of any further combinatorial problems (see [2, Appendix 1]).

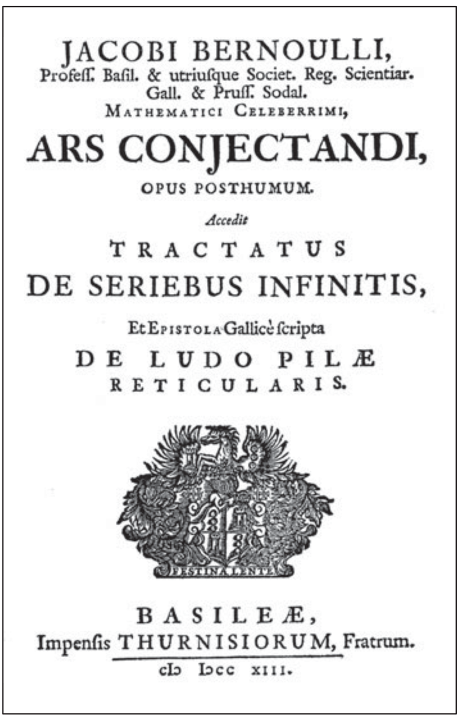
It is interesting to note that Pascal seemed averse to using any argument involving $n!$ as the number of permutations of n different things; it is almost as though he felt that he had a mission to develop the theory of combinations without using it, and he certainly never gave it. When his friend M. de Gaignières challenged him to find an explanation of the multiplicative formula, he said that because of the difficulty he thought it proper to leave the demonstration to him; 'however, thanks to the Arithmetical Triangle, an easy way was opened up', and he pointed out the identity in one of the 'consequences' of his book.

De Montmort and Jacob Bernoulli

It is from Pierre de Montmort that Pascal probably acquired the credit for the arithmetical triangle, for de Montmort's book *Essay d'Analyse sur les Jeux de Hazard* (Essay on the Analysis of Games of Chance), whose principal edition is dated 1713, starts with a seventy-two page *Traité des Combinaisons* (Treatise on Combinations) which relies on Pascal's *Traité*. In his introduction, de Montmort wrote that

> Pascal has proceeded furthest, as is clear from his treatise *The Arithmetical Triangle*, which is full of observations and discoveries on the figurate numbers of which I believe him to be the originator, since he does not cite any other person.

Already in its first edition in 1708, de Montmort had given the combinatorial explanation for $C(n, r)$, which we earlier attributed to ben Gerson. Unlike Pascal, de Montmort proceeded by revealing patterns of enumeration that enabled him to identify numbers of combinations with the figurate numbers (and hence the arithmetical triangle), but much of his *Traité* was devoted to combinatorial problems that cannot easily be solved in terms of the numbers in the triangle. One problem that did lead to the triangle was Tartaglia's dice problem, which de Montmort solved by enumeration, describing it as 'curious enough, it seems to me', and which is the same problem as finding the number of terms in the power of a multinomial.



JACOBI BERNOULLI,
Profeſſ. Baſil. & utriuſque Societ. Reg. Scientiar.
Gall. & Pruſſ. Sodal.
MATHEMATICI CELEBERRIMI,

ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè ſcripta
DE LUDO PILÆ
RETICULARIS.

FESTINA LENTE

BASILEÆ,
Impenſis THURNISIORUM, Fratrum.
cIↃ IↃcc xiii.

Title page of Jacob Bernoulli's *Ars Conjectandi*.

Part Two of Jacob Bernoulli's *Ars Conjectandi* (Art of Conjecturing) of 1713 is *Doctrina de Permutationibus & Combinationibus* (A Treatise on Permutations and Combinations), which was written in apparent ignorance of Pascal's *Traité*. The arithmetical triangle makes its appearance in Chapter III as the solution to the number of combinations, enumeration being along the same lines as de Montmort's. Bernoulli waxed lyrical about it:

> This Table has truly exceptional and admirable properties; for besides concealing within itself the mysteries of combinations, as we have seen, it is known by those expert in the higher parts of mathematics also to hold the foremost secrets of the whole of the rest of the subject.

He then listed twelve 'wonderful properties' of the triangle, rather as Pascal had done, concluding with a proof of the multiplicative formula which is longer and less elegant than Pascal's. Like Pascal, he failed to provide a direct combinatorial proof.

The outstanding contribution of Part Two is the derivation of the coefficients of the polynomials for the sums of the powers of the integers, with which Bernoulli ended Chapter III (see Chapter 6). Although this relies on the figurate numbers, the argument is not combinatorial; further information can be found in [2, Ch. 10 and Appendix 3].

In the remaining chapters of Part Two of his *Ars Conjectandi*, Bernoulli dealt with a number of combinatorial problems in ways by now familiar, but mention may be made of the arithmetical triangle that appears in Chapter IV in connection with the problem of points. Bernoulli here added nothing to Pascal's solution. In Chapter V there is another arithmetical triangle, this time presenting the number of combinations of r things from n different kinds of things, repeats being allowed. The result was obtained by a clever systematic enumeration that demonstrates the applicability of the addition rule for figurate numbers, and Bernoulli then gave a combinatorial explanation for this rule. The problem is identical to that solved by Tartaglia, which we discussed earlier. In Chapter VIII Bernoulli remarked, like de Montmort, that it is also the solution to the problem of finding the number of terms in the power of a multinomial:

> It is proper here to note the peculiar sympathy between combinations and powers of multinomials.

In the introduction to Part Two, Bernoulli mentioned the names of van Schooten, Leibniz, Wallis, and Prestet as having preceded him. Amongst these authors we remark only that Leibniz, in his youthful *Dissertatio de Arte Combinatoria* of 1666, repeated the combinatorial argument that (unknown to him) Pascal had already used in his *Traité* to obtain the addition relation in the arithmetical triangle. Leibniz presented a table of the triangle up to $n = 12$; the many further contributions he made to combinatorial theory were discussed in Chapter 6.

From binomials to multinomials

The arithmetical triangle naturally generalizes to more dimensions, corresponding to the generalization of binomial coefficients to multinomial coefficients. The rule is that the number of permutations of a things of one kind, b of a second kind, c of a third kind, and so on, n things all told, is equal to

$$\frac{n!}{a!\,b!\,c!\cdots}.$$

This rule first appeared in the West in the work of Mersenne in 1636, and was later explained by Wallis in 1685. As we mentioned earlier, Bhāskara had already given it in the East in his *Līlāvatī*. On 16 May 1695 Leibniz wrote to Johann Bernoulli announcing a 'wonderful rule' for the coefficients of the powers of a multinomial, to which Bernoulli replied on 8 June giving the above formula and adding

> It would be a pleasure to see your rule and it would be well to test whether they agree; yours is possibly simpler.

(In fact it was slightly more complicated.) De Moivre published this multinomial coefficient in 1698, observing simply that it gives the number of permutations of the elements making up a given term.

Notes

1. D. E. Knuth, *The Art of Computer Programming*, Vol. I, Fundamental Algorithms (2nd edn.), Addison-Wesley (1973), 53.
2. For a full history of the arithmetical triangle and its influence in the development of mathematics in general, and for references to all the writers mentioned in this chapter, see A. W. F. Edwards, *Pascal's Arithmetical Triangle*, Johns Hopkins University

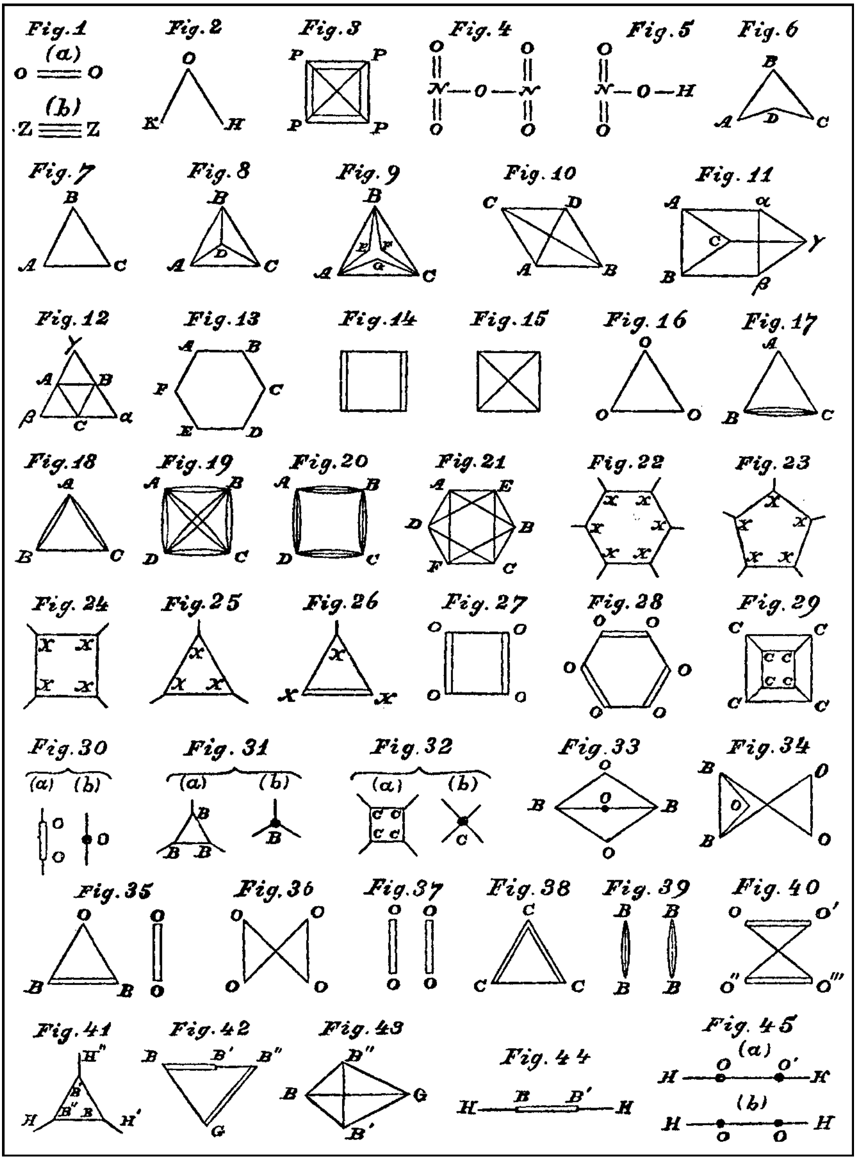
Press (2002). This second edition contains an epilogue and a third appendix not present in the first edition (Charles Griffin & Co., Ltd, and Oxford University Press (1987)).

The epilogue lists a number of books and articles that had appeared since 1987, or which were earlier than 1987 but originally overlooked. The most significant oversight was the excellent translation of Pascal's *Traité* and his correspondence with Fermat in *Great Books of the Western World* 33: *The Provincial Letters, Pensées, Scientific Treatises by Blaise Pascal* (Encyclopaedia Britannica, 1955, 1963). Cambridge University Library has digitized one of its copies of the *Traité* at www.lib.cam.ac.uk/RareBooks/PascalTraite/.

Appendix 3 is a commentary on the Introduction and Chapters I–III of Bernoulli's *Ars Conjectandi*, Part Two: The theory of permutations and combinations, and in particular contains descriptions of some of the proofs in Bernoulli's Chapter III. In 2006 the first complete translation into English of *Ars Conjectandi* was published: *J. Bernoulli, The Art of Conjecturing*, together with *Letter to a Friend on Sets in Court Tennis*, translated with an introduction and notes by Edith Dudley Sylla (Johns Hopkins University Press). Some remarks relevant to combinatorics are in A. W. F. Edwards's review of this translation in *The Mathematical Intelligencer* 29 (2007), 70–2.

PART III

MODERN COMBINATORICS



J. J. Sylvester's chemical trees.

CHAPTER 8

Early graph theory

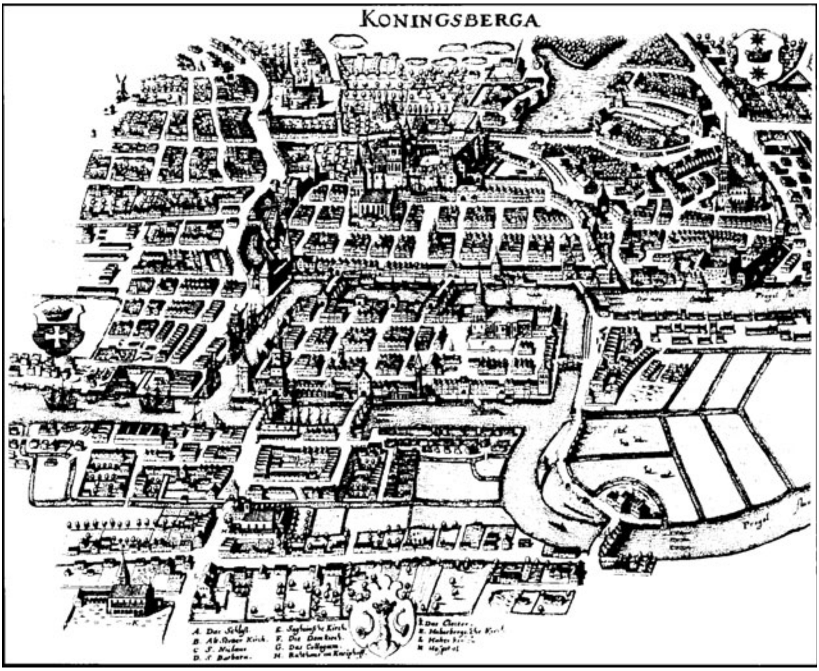
ROBIN WILSON

The origins of graph theory are humble, even frivolous. Whereas many branches of mathematics were motivated by fundamental problems of calculation, motion, and measurement, the problems which led to the development of graph theory were often little more than puzzles, designed to test the ingenuity rather than to stimulate the imagination. But despite the apparent triviality of such puzzles, they captured the interest of mathematicians, with the result that graph theory has become a subject rich in theoretical results of a surprising variety and depth.

So begins the book *Graph Theory 1736–1936* [3], which chronicles the history of graph theory from Euler's treatment of the Königsberg bridges problem in the 1730s to the explosion of activity in the area in the 20th century. In this chapter, and in Chapter 14, we present this story [1].

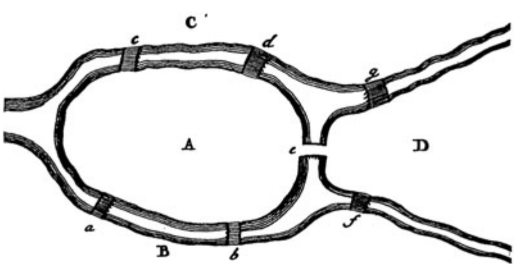
The Königsberg bridges problem

On 26 August 1735 Leonhard Euler lectured on 'the solution of a problem relating to the geometry of position' to his colleagues at the Academy of Sciences in St Petersburg, where he had worked since 1727. In his account, Euler presented his solution of the *Königsberg bridges problem*, which he believed to be widely known, and which asked whether it was possible to find a route crossing each of the seven bridges of Königsberg once and only once. More generally, given any division of a river into branches and any arrangement of bridges, is there a general method for determining whether such a route is possible?



A 17th-century map of Königsberg.

In 1736 Euler communicated his solution to his friend Carl Ehler, Mayor of Danzig, and to the Italian mathematician Giovanni Marinoni (see [18], [37]). He also wrote up his solution for publication in the *Commentarii Academiae Scientiarum Imperialis Petropolitanae* under the title 'Solutio problematis ad geometriam situs pertinentis' (On the solution of a problem pertaining to the geometry of position) [11]. Although dated 1736, it did not appear until 1741, and was later republished in the new edition of the *Commentarii* (*Nova Acta Commentarii* . . .), which appeared in 1752.



Königsberg, from Euler's 1736 paper.

Euler's paper is divided into twenty-one numbered paragraphs, of which paragraphs 2–9 show the impossibility of solving the Königsberg bridges problem and the rest are concerned with the general situation. He first described the problem as relating to the *geometry of position*, a branch of mathematics first mentioned by Leibniz and concerned with aspects of position rather than the calculation of magnitudes; some interpretations that have been put on this phrase are discussed by Pont [32]. Euler then reformulated the problem as one of trying to

Find a sequence of eight letters A, B, C, or D (the land areas) *such that the pairs AB and AC are adjacent twice* (corresponding to the two bridges between *A* and *B* and between *A* and *C*)*, and the pairs AD, BD, and CD are adjacent just once* (corresponding to the other bridges).

He then showed why this is impossible.

In discussing the general problem, Euler observed that

The numbers of bridges written next to the letters *A*, *B*, *C*, etc. together add up to twice the total number of bridges. The reason for this is that, in the calculation where every bridge leading to a given area is counted, each bridge is counted twice, once for each of the two areas that it joins.

This is the earliest statement of what graph-theorists now call the *handshaking lemma*. The paper continues with Euler's main conclusions:

If there are more than two areas to which an odd number of bridges lead, then such a journey is impossible.

If, however, the number of bridges is odd for exactly two areas, then the journey is possible only if it starts in either of these two areas.

If, finally, there are no areas to which an odd number of bridges lead, then the required journey can be accomplished starting from any area.

Euler then remarked:

When it has been determined that such a jouney can be made, one still has to find how it should be arranged. For this I use the following rule … .

However, he did not prove that his rule can always be carried out. A valid demonstration did not appear until a related result was proved by Carl Hierholzer [17] in 1873. Hierholzer's discussion was given in the language of diagram tracing, to which we now turn.

Diagram-tracing puzzles

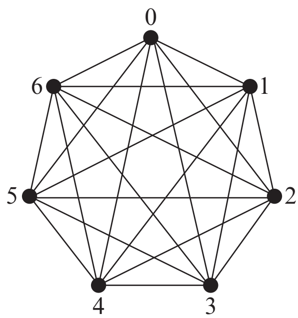
In 1809 the French mathematician Louis Poinsot wrote a memoir on polygons and polyhedra [29] in which he described the four non-convex regular polyhedra and posed several geometrical problems, including the following:

Given some points situated at random in space, it is required to arrange a single flexible thread uniting them two by two in all possible ways, so that finally the two ends of the thread join up, and so that the total length is equal to the sum of all the mutual distances.

For example, we can arrange a thread joining the seven points 0 to 6 in the order

$$0 \to 1 \to 2 \to 3 \to 4 \to 5 \to 6 \to 0 \to 2 \to 4 \to 6 \to 1 \to 3 \to 5 \to$$
$$0 \to 3 \to 6 \to 2 \to 5 \to 1 \to 4 \to 0.$$

Poinsot noted that a solution is possible only for an odd number of points, and gave an ingenious method for joining the points in each such case.



Joining seven points.

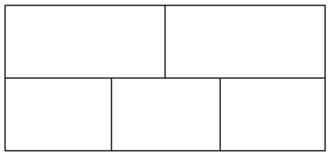
In fact, there are millions of solutions, as was later observed by M. Reiss [35] in the context of determining the number of ways that one can lay out a complete set of dominoes in a ring; the above ordering corresponds to the ring of twenty-one dominoes

0–1, 1–2, 2–3, 3–4, 4–5, 5–6, 6–0, 0–2, 2–4, 4–6, 6–1,

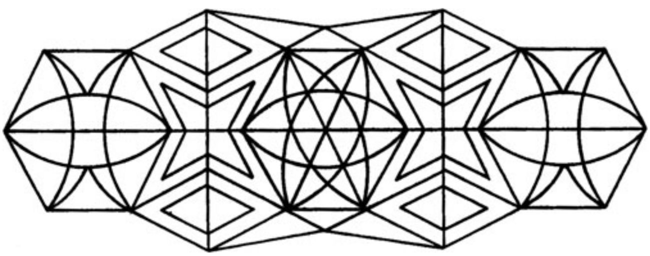
1–3, 3–5, 5–0, 0–3, 3–6, 6–2, 2–5, 5–1, 1–4, 4–0,

to which the seven doubles (such as 2–2) are then added.

Puzzles that require one to draw a given diagram in the smallest possible number of connected strokes, without any overlapping, have been of interest for many hundreds of years – see, for example, some early African examples in Ascher [2]. In particular, it was observed in 1844 that four strokes are needed to draw the following diagram.



In 1847 Johann Benedict Listing wrote a short treatise entitled *Vorstudien zur Topologie* [24], in which he investigated a number of non-metrical geometrical problems and discussed the solution of diagram-tracing puzzles; these included the above example and a much more complicated diagram (below) that can be drawn in a single stroke. His treatise is noteworthy for being the first place that the word 'topology' appeared in print. Listing had first coined the word in 1836 in a letter to his former schoolteacher.



Listing's diagram.

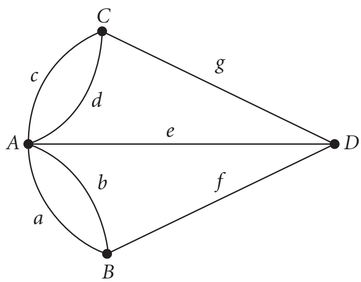
As mentioned above, Hierholzer gave an account of the theory of diagram-tracing puzzles, proving in particular that:

If a line-system can be traversed in one path without any section of line being traversed more than once, then the number of odd nodes is either zero or two.

He also proved the converse result:

If a connected line-system has either no odd node or two odd nodes, then the system can be traversed in one path.

The precise connection between Euler's bridge-crossing problems and the tracing of diagrams was not made until the end of the 19th century. Euler's discussion of such problems had been popularized through a French translation of E. Coupy [8] that included an application to the bridges over the River Seine, and by a lengthy account in Volume 1 of E. Lucas's *Récréations Mathématiques* [26], but it seems to have been W. W. Rouse Ball [36] who first represented the four land areas by points and the bridges by lines joining the appropriate pairs of points, thereby producing the well-known four-point diagram. Euler never drew such a picture or discussed the Königsberg bridges problem in these terms.



The graph of the Königsberg bridges.

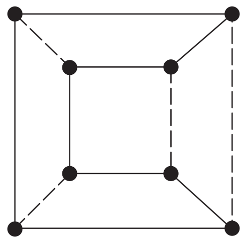
Such a diagram is now called a *connected graph*, the points are *vertices*, the lines are *edges*, and the number of edges appearing at a vertex is the *degree* of that vertex; thus, the above graph has three vertices of degree 3 and one vertex of degree 5. It follows from the above results that:

A connected graph has a path that includes each edge just once if and only if there are exactly 0 *or* 2 *vertices of odd degree*.

When there are no vertices of odd degree, the graph is called an *Eulerian graph*, even though the concept of such a graph did not make its first appearance until many years after the solution by Euler that inspired it.

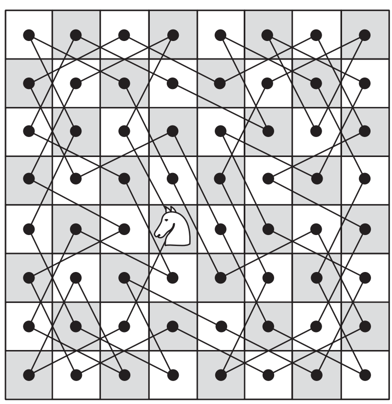
Hamiltonian graphs

A type of graph problem that is superficially similar to the Eulerian problems described above is that of *finding a cycle that passes just once through each vertex*, rather than just once along each edge; for example, if we are given the graph of the cube, then it is impossible to cover each *edge* just once, because there are eight vertices of degree 3, but we can find a cycle (shown below with solid lines) passing through each *vertex* just once. Such graphs are now called *Hamiltonian graphs*, and the cycles are *Hamiltonian cycles*, although, as we shall see, this is perhaps not the most appropriate name for them.



A Hamiltonian cycle on a cube.

An early example of a Hamiltonian cycle problem is the celebrated *knight's tour problem*. The problem is to find a succession of knight's moves on a chessboard visiting each of the sixty-four squares just once and returning to the starting point. The connection with Hamiltonian graphs may be seen by regarding the squares as vertices of a graph, and joining two squares whenever they are connected by a single knight's move.



A knight's tour on a chessboard.

Solutions of the knight's tour problem have been known for many hundreds of years, including solutions by de Montmort and de Moivre in the 17th century, but it was not until the mid 18th century that the problem was subjected to systematic mathematical analysis, by Leonhard Euler [13]; Euler showed in particular that no solution is possible for the analogous problem on a chessboard with an odd number of squares. Shortly afterwards, A.-T. Vandermonde [42] analysed the problem, referring to Euler's solution as follows:

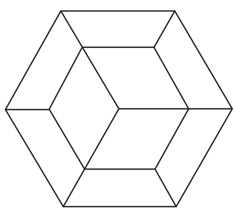
> whereas that great geometer presupposes that one has a chessboard to hand, I have reduced the problem to simple arithmetic.

Many mathematicians have since attempted to generalize the problem to other types of board or to find solutions that satisfy extra conditions; for example, Major Carl von Jaenisch [19] wrote a three-volume account of the knight's tour problem, and included an ingenious solution in which successive numbering of the squares in a knight's tour yields a semi-magic square in which the entries in each row or column add up to 260.

In 1855 the Royal Society of London received a paper by the Revd Thomas Penyngton Kirkman [21] that asked:

For which polyhedra can one find a cycle passing through all the vertices just once?

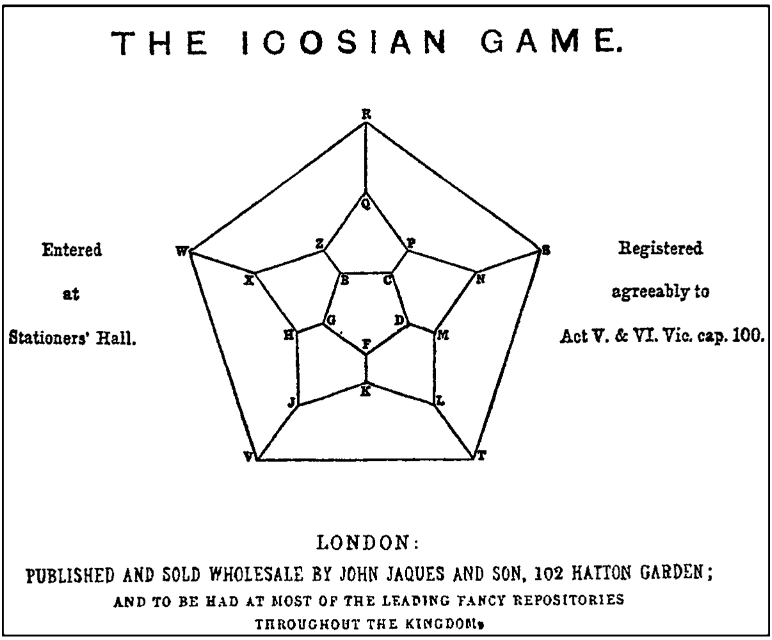
– for example, a cube (in flattened form) has the cycle given earlier. Kirkman claimed to have a sufficient condition for the existence of such a cycle, but his reasoning was faulty. However, he did explain why any polyhedron with even-sided faces and an odd number of vertices can have no such cycle, and gave as an example the polyhedron obtained by 'cutting in two the cell of a bee'.



Kirkman's 'cell of a bee'.

A polyhedron is *cubic* if exactly three faces meet at each vertex. In 1884 the natural philosopher P. G. Tait asserted that every cubic polyhedron has a cycle passing through every vertex. If true, this assertion would have yielded a simple proof of the four-colour theorem (see later), but it was eventually disproved by W. T. Tutte [41], who in 1946 presented a cubic polyhedron with forty-six vertices and no Hamiltonian cycle.

Another mathematician who was intrigued with cycles on polyhedra was Sir William Rowan Hamilton (see [15]). Arising from his work on quaternions and non-commutative algebra, Hamilton was led to the *icosian calculus*, in which he considered cycles of faces on an icosahedron – or, equivalently, cycles of vertices on a dodecahedron. Hamiltonian subsequently invented the *icosian game*, based on a dodecahedron whose vertices were labelled with the twenty consonants *B* to *Z*, standing for Brussels, Canton, . . . , Zanzibar. The object of the game was to find cycles through all the vertices according to certain specified instructions. Hamilton sold the game for £25 to a games manufacturer who marketed it under the name 'A voyage round the world'; not surprisingly, it was not a commercial success.



Hamilton's icosian game.

Because of Hamilton's prominence, his name has become associated with such cycle problems and with the corresponding Hamiltonian graphs, even though Kirkman considered these problems in greater generality and had preceded him by a few months. Unlike the Eulerian problem, no necessary and sufficient condition has been discovered for the existence of a Hamiltonian cycle in a general graph.

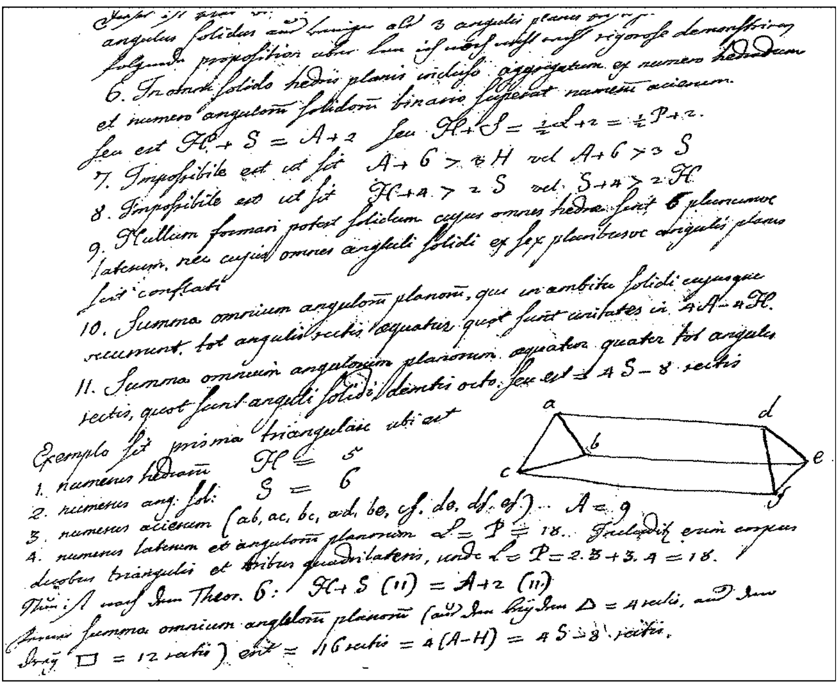
Euler's polyhedron formula

In this section we investigate the origins of the polyhedron formula for both polyhedra and planar graphs, and show how one of its generalizations led to the work of Listing and, ultimately, Poincaré.

Although the Greeks were familiar with the five regular solids (the tetrahedron, cube or hexahedron, octahedron, dodecahedron, and icosahedron) and several other polyhedra, there is no evidence that they knew the simple formula relating the numbers of vertices, edges, and faces of such a polyhedron – namely,

$$(\textit{number of faces}) + (\textit{number of vertices}) = (\textit{number of edges}) + 2;$$

for example, a cube has six faces, eight vertices, and twelve edges, and $6 + 8 = 12 + 2$. In the 17th century, René Descartes also missed the formula: he obtained a formula for the sum of the angles in all the faces of a polyhedron, from which the above result can be deduced, but he never made the deduction. It was Euler, in a letter to Christian Goldbach in 1750, who introduced the concept of an edge and stated the above result.


6. In omni solido hedris planis incluso aggregatum ex numero hedrarum et numero angulorum solidorum binario superat numerum acierum. seu est H + S = A + 2 seu H + S = ½L + 2 = ½P + 2.
7. Impossibile est ut sit A + 6 > 3H vel A + 6 > 3S
8. Impossibile est ut sit H + 4 > 2S vel S + 4 > 2H
9. Nullum formari potest solidum cujus omnes hedrae sint 6 pluribusve lateribus, nec cujus omnes anguli solidi ex sex pluribusve angulis planis sint conflati
10. Summa omnium angulorum planorum, qui in ambitu solidi cujusque occurrunt, tot angulis rectis aequatur quot sunt unitates in 4A − 4H.
11. Summa omnium angulorum planorum aequatur quater tot angulis rectis, quot sunt anguli solidi demtis octo. seu est = 4S − 8 rectis
Exempli loco sit prisma triangulare, ubi est
1. numerus hedrarum H = 5
2. numerus ang. sol. S = 6
3. numerus acierum (ab, ac, bc, ad, be, cf, de, df, ef) A = 9
4. numerus laterum et angulorum planorum L = P = 18.

Part of Euler's letter to Goldbach, presenting his 'polyhedron formula'.

In his letter Euler considered a solid polyhedron and obtained various results concerning equalities and inequalities about the numbers of faces, solid angles (vertices), and joints where two faces come together (edges). In particular, denoting them respectively by H (*hedrae*), S (*anguli solidi*), and A (*acies*), he asserted that:

> 6. In every solid enclosed by plane faces the aggregate of the number of faces and the number of solid angles exceeds by two the number of edges, or $H + S = A + 2 \ldots$
>
> 11. The sum of all plane angles is equal to four times as many right angles as there are solid angles, less eight, that is $= 4S - 8$ right angles ...
>
> I find it surprising that these general results in solid geometry have not previously been noticed by anyone, so far as I am aware; and furthermore, that the important ones, Theorems 6 and 11, are so difficult that I have not yet been able to prove them in a satisfactory way.

Euler verified these results for several families of polyhedra and two years later presented a dissection proof [12], but this was deficient. The first valid proof was a metrical one, given in 1794 by A.-M. Legendre [22].

Euler's formula also holds for any connected *planar graph*, such as the map obtained by stereographically projecting a polyhedron onto a plane, provided that we remember to include the 'infinite' (unbounded) face. In 1813 Augustin-Louis Cauchy [4] used a triangulation process to give topological proofs of both versions of Euler's formula, and deduced that there are only four regular non-convex polyhedra, as Poinsot had predicted.

Around the same time Simon-Antoine-Jean Lhuilier [23] gave a topological proof that there are only five regular convex polyhedra, and anticipated the idea of duality by remarking that four of them occur in reciprocal pairs (cube–octahedron and dodecahedron–icosahedron), while the tetrahedron is self-dual. He also found three types of polyhedra for which Euler's formula fails – those with an interior cavity, those with indentations in their faces, and ring-shaped polyhedra drawn on a torus (that is, polyhedra containing a 'tunnel'). For ring-shaped polyhedra, he obtained the formula

$$(\textit{number of vertices}) - (\textit{number of edges}) + (\textit{number of faces}) = 0.$$

He then extended his discussion to prove that if g is the number of 'tunnels' in a surface on which a polyhedral map is drawn, then

$$(\textit{number of vertices}) - (\textit{number of edges}) + (\textit{number of faces}) = 2 - 2g.$$

The number g is now called the *genus* of the surface, and the quantity $2 - 2g$ is its *Euler characteristic*; these numbers depend only on the surface on which the polyhedron is embedded, and not on the map itself.

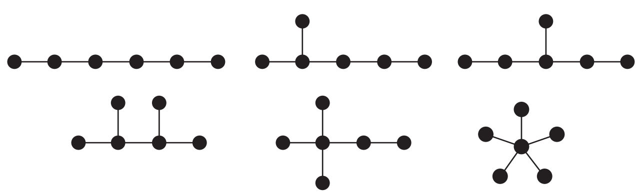
Lhuilier's results were the starting point for an extensive investigation by Listing [25], entitled *Der Census räumliche Complexe* (The Census of Spatial Complexes). These complexes are built up from simpler pieces, and Listing studied the question of how their topological properties are related to the above versions of Euler's formula.

Listing's ideas proved to be influential in the subsequent development of topology, and were soon taken up by other mathematicians. In particular, Henri Poincaré developed them in his papers of 1895 to 1904 that laid the foundations of algebraic topology. Like Listing, Poincaré developed a method for constructing complexes from basic 'cells', such as 0-cells (vertices) and 1-cells (edges). In order to fit these cells together, he adapted a technique of Gustav Kirchhoff from the theory of electrical networks, replacing sets of linear equations by matrices. These matrices could then be studied from an algebraic point of view.

Poincaré's work was an instant success, and appeared in M. Dehn and P. Heegaard's article (1907) on *analysis situs* (the analysis of position) in the *Encyklopädie der Mathematischen Wissenschaften*. His ideas were subsequently developed further by Oswald Veblen in a series of colloquium lectures for the American Mathematical Society on analysis situs; these lectures were delivered in 1916 and published in book form six years later [43]. The subject subsequently became known as topology.

Trees

We are all familiar with the idea of a family tree. Mathematically, a *tree* is a connected graph with no cycles; the following figure illustrates the possible trees with six vertices.

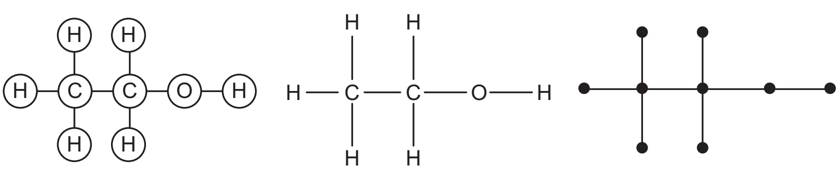


The trees with six vertices.

The concept of a tree appeared implicitly in 1847 in the work of Kirchhoff, who used them in the calculation of currents in an electrical network. Here, however, our concern is mainly with the problem of enumerating certain types of chemical molecule. Such problems can be reduced to the counting of trees and were investigated by Arthur Cayley and James Joseph Sylvester. We outline their contributions and indicate how their ideas were developed in the first half of the 20th century.

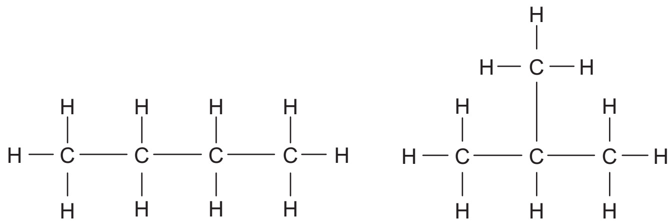
By 1850 it was already known that chemical elements combine in fixed proportions, and chemical formulas such as H_2O (water) and C_2H_5OH (ethanol) were well established. But it was not understood exactly how the various elements combine to form these substances. The breakthrough occurred in the 1850s when August Kekulé in Germany, Edward Frankland in England, A. M. Butlerov in Russia, and A. S. Couper in Scotland proposed what is now the theory of *valency*; in this theory, each atom has several bonds by which it is linked to other atoms: *carbon atoms have four bonds, oxygen atoms have two, and hydrogen atoms have one*.

As the idea of valency became established, it became increasingly necessary for chemists to find a method for representing molecules diagrammatically. Various people tried and failed, including some of those mentioned above, and it was not until the 1860s that Alexander Crum Brown [9] proposed what is essentially the form we use today. In his system, each atom is represented by a circled letter and the bonds are indicated by lines joining the circles. The following figure shows Crum Brown's representation of ethanol, the present-day representation with the circles omitted, and the associated chemical tree with vertices corresponding to atoms and edges representing bonds.



Various representations of ethanol.

Crum Brown's 'graphic notation', as it came to be called, was quickly adopted by Frankland, who used it in his *Lecture Notes for Chemical Students* [14]. Its great advantage was that its use explained, for the first time, the phenomenon of *isomerism*, whereby there can exist pairs of molecules (*isomers*) with the same chemical formula but different chemical properties. The following figure shows a pair of isomers, each with chemical formula C_4H_{10}; note how the atoms are arranged in different ways inside the molecules.



n-butane and 2-methylpropane.

This idea leads naturally to problems of *isomer enumeration*, in which we determine the number of different molecules with a given chemical formula. The most celebrated of these problems is that of enumerating the alkanes (paraffins), which have chemical formula C_nH_{2n+2}; the following table gives the numbers of such molecules with up to eight carbon atoms:

Formula	CH_4	C_2H_6	C_3H_8	C_4H_{10}	C_5H_{12}	C_6H_{14}	C_7H_{16}	C_8H_{18}
Number	1	1	1	2	3	5	9	18

In 1874 Cayley observed that the diagrams corresponding to the alkanes all have a tree-like structure, and that removing the hydrogen atoms yields a tree in which each vertex has degree 1, 2, 3, or 4; thus, the problem of enumerating such isomers is the same as that of counting trees with this property.

Cayley had been interested in tree-counting problems for some time. In 1857, while trying to solve a problem inspired by Sylvester related to the differential calculus, he wrote a paper [5] in which he enumerated *rooted trees* – that is, trees in which one particular vertex has been singled out as the 'root' of the tree. As described in greater detail in Chapter 12, Cayley's method was to remove the root, thereby obtaining a number of smaller rooted trees.

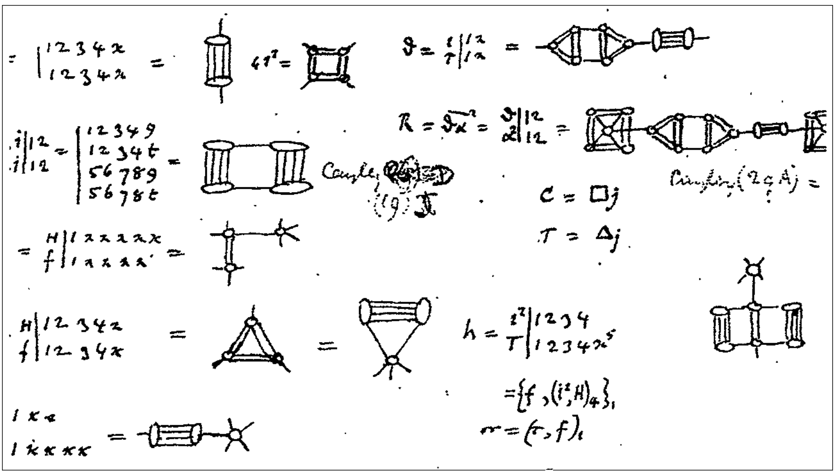
At around the same time that Cayley was enumerating isomers, his friends Sylvester and William Clifford were trying to establish a link between the study of chemical molecules and the algebraic topic of *invariant theory*. Each chemical atom was to be compared with a 'binary quantic', a homogeneous expression in two variables such as

$$ax^3 + 3bx^2y + 3cxy^2 + dy^3,$$

and a chemical substance composed of atoms of various valencies was to be compared with an 'invariant' of a system of binary quantics of the corresponding degrees. Indeed, profoundly influenced by Frankland's *Lecture Notes*, Sylvester was later to write [39]:

> The more I study Dr Frankland's wonderfully beautiful little treatise the more deeply I become impressed with the harmony or homology . . . which exists between the chemical and algebraical theories. In travelling my eye up and down the illustrated pages of "the Notes," I feel as Aladdin must have done in walking in the garden where every tree was laden with precious stones . . .

Both Cayley and Sylvester had made important contributions to the theory of invariants, and Sylvester and Clifford tried to introduce the 'graphic notation' of chemistry into the subject; indeed, our use of the word *graph* for such a diagram arose from one of Sylvester's papers [38] in this area.

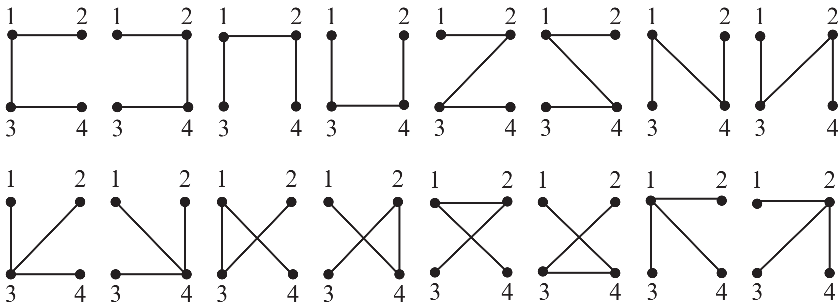


Some of Clifford's drawings of binary quantics.

Unfortunately, the 'chemico-algebraic' ideas of Sylvester and Clifford proved to be less useful than their originators had hoped – two largely unrelated ideas linked by a notation that was only superficially similar. Invariant theory quickly became submerged in the work of David Hilbert and others, while the theory of graphs increasingly took on a life of its own (see Chapter 14). Perhaps Sylvester feared this all along; in a letter to Simon Newcomb, he nervously admitted that:

I feel anxious as to how it will be received as it will be thought by many strained and over-fanciful. It is more a 'reverie' than a regular mathematical paper . . . [Nevertheless,] it may at the worst serve to suggest to chemists and Algebraists that they may have something to learn from each other.

In 1889 Cayley tackled another tree-counting problem – that of determining the number $t(n)$ of labelled trees with n vertices; for example, if $n = 4$, the number of such trees is 16. Unlike the earlier problems we considered, this one has a very simple answer – namely, $t(n) = n^{n-2}$. Cayley [7] stated this result and demonstrated it for $n = 6$, but Heinz Prüfer [33] was the first to publish a complete proof, in 1918.



The sixteen labelled trees with four vertices.

It was not until the 1920s and 1930s that any substantial theoretical progress was made in the counting of chemical molecules. In 1927 J. H. Redfield [34] produced a paper that foreshadowed the later work of George Pólya, but it was written in obscure language and overlooked for many years. Shortly afterwards, A. C. Lunn and J. K. Senior [27] recognized that the theory of permutation groups was appropriate to the enumeration of isomers, and their ideas were considerably developed in a fundamental paper of Pólya [30] (translated in [31]), in which the classical method of generating functions is combined with the idea of a permutation group in order to enumerate graphs and molecules and many other configurations arising in mathematics. Details of the work of Redfield and Pólya are given in Chapter 12.

The four-colour problem

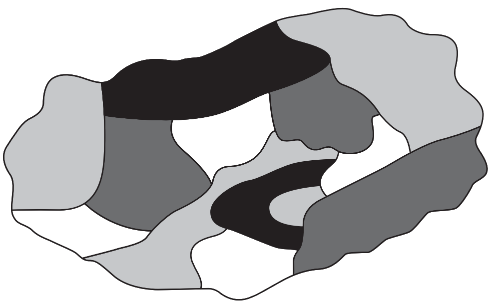
The earliest known reference to the 'four-colour problem' on the colouring of maps occurs in a letter dated 23 October 1852, from Augustus De Morgan to Sir

William Rowan Hamilton. In this letter (see [44]), De Morgan described how one of his students had asked him whether every map can be coloured with only four colours:

A student of mine asked me today to give him a reason for a fact which I did not know was a fact – and do not yet. He says that if a figure be anyhow divided and the compartments differently coloured so that figures with any portion of common boundary *line* are differently coloured – four colours may be wanted, but not more . . . Query cannot a necessity for five or more be invented.

The student was later identified as Frederick Guthrie, who claimed that the problem was due to his brother Francis; the latter had formulated it while colouring the counties of a map of England.

In his letter, De Morgan observed that four colours are needed for some maps; for example, if there are four neighbouring countries, then each country must be coloured differently from its neighbours. But four colours may be needed even if four neighbouring countries do not appear, as the following example shows:



A map that needs four colours.

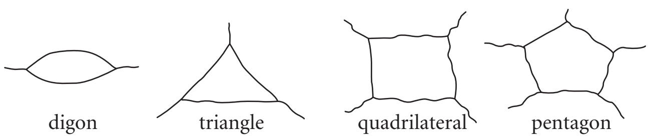
De Morgan quickly became intrigued by the problem and communicated it to several other mathematicians, so that it soon became part of mathematical folklore. In 1860 he stated it, in rather obscure terms, in an unsigned book review [10] in the *Athenaeum*, a scientific and literary journal. For many years this was believed to be the first printed reference to the problem, but an earlier *Athenaeum* reference, dated 1854 and signed 'F.G.', was recently found by Brendan McKay [28]. De Morgan's review was read in the USA by the logician and philosopher C. S. Peirce, who subsequently presented an attempted proof to a mathematical society at Harvard University.

It was not until after De Morgan's death in 1871 that any progress was made in solving the four-colour problem. On 13 June 1878, at a meeting of the London

Mathematical Society, Cayley enquired whether the problem had been solved, and soon afterwards wrote a short paper [6] for the Royal Geographical Society in which he attempted to explain in simple terms where the difficulties lie. He also proved that one can make the simplifying assumption that exactly three countries meet at each point – that is, the map is *cubic*.

In 1879 there appeared one of the most famous fallacious proofs in mathematics. Its author was Alfred Bray Kempe, a London barrister who had studied with Cayley at Cambridge, had attended the London Mathematical Society meeting, and had become well known for his work on linkages. On learning of this proof, Cayley suggested that Kempe submit it to the *American Journal of Mathematics*, newly founded and edited by Sylvester.

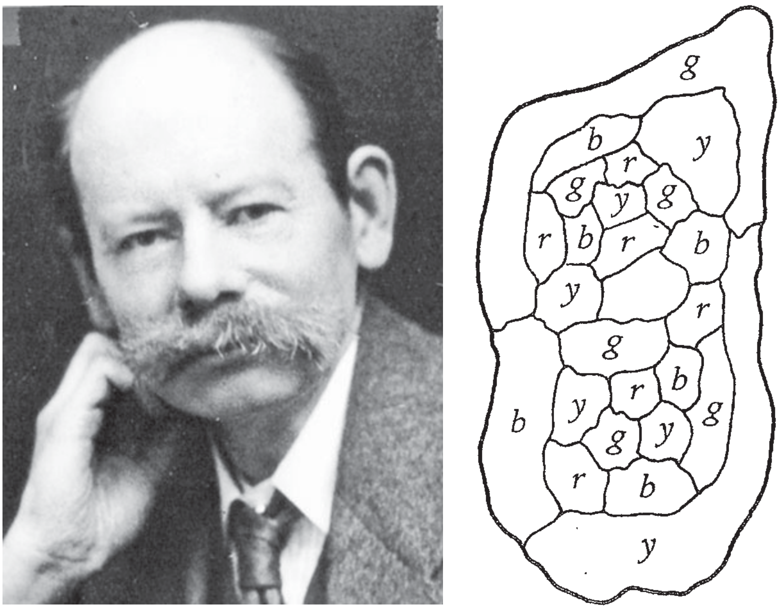
Although Kempe's paper [20] contained a fatal flaw, it included some important ideas that were to feature in many subsequent attempts on the problem. His proof was in two parts. He first showed, using Euler's polyhedron formula, that every map necessarily contains a digon (a two-sided country), a triangle, a quadrilateral, or a pentagon, as shown below.



Kempe then took each of these configurations in turn and showed that any colouring containing it can be extended to the whole map. Now, it is simple to prove that this process can be carried out for the digon and triangle. To prove that it can also be done for a quadrilateral, Kempe looked at a two-coloured piece of the map – for example, the part of the map containing countries coloured red and green – and he showed how to interchange the colours so as to enable the original colouring to be extended to the whole map as required. To prove that the above process can also be carried out for a pentagon, Kempe did it twice, making *two* colour interchanges simultaneously. Since all possible cases had been considered, the proof was believed complete.

Kempe's argument was greeted with enthusiasm, and he published two further papers indicating various simplifications. In 1880 Tait [40] reformulated the result in terms of the colouring of 'boundary' *edges* (rather than countries), believing that such considerations would simplify the proof still further. The headmaster of a famous school set the problem as a challenge problem to his pupils, Frederick Temple (Bishop of London, later Archbishop of Canterbury) produced a 'proof' during a lengthy meeting, and Lewis Carroll reformulated the problem as a game between two players.

In 1890 Percy Heawood, who had learned of the problem while a student at Oxford University, published a paper [16] that pointed out Kempe's error. Furthermore, Heawood gave a specific example to show that, whereas one colour interchange is always permissible, we cannot carry out two interchanges at the same time; thus, Kempe's treatment of the pentagon was deficient.



Percy Heawood and his example.

Heawood managed to salvage enough from Kempe's argument to prove that *every map can be coloured with five colours* (itself a remarkable result), but he was unable to fill the gap. How this was eventually done is described in Chapter 14 and in [44].

References

1. An earlier version of this chapter appears in N. L. Biggs, E. K. Lloyd, and R. J. Wilson, The history of combinatorics, *Handbook of Combinatorics* (ed. R. Graham *et al.*), Elsevier (1995), Chapter 44.

2. M. Ascher, *Ethnomathematics: A Multicultural View of Mathematical Ideas*, Brooks & Cole (1991).
3. N. L. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory 1736–1936*, Clarendon Press, Oxford (1976) (reissued in paperback in 1998).
4. A.-L. Cauchy, Recherches sur les polyèdres – premier mèmoire, *J. Ecole Polytech*. 9 (Cah. 16) (1813), 68–86 = *Oeuvres (2)*, Vol. 1, 7–25.
5. A. Cayley, On the theory of the analytical forms called trees, *Philos. Mag. (4)* 13 (1857), 172–6 = *Math. Papers*, Vol. 3, 242–6.
6. A. Cayley, On the colouring of maps, *Proc. Roy. Geog. Soc. (New Ser.)* 1 (1879), 259–61 = *Math. Papers*, Vol. 11, 7–8.
7. A. Cayley, A theorem on trees, *Quart. J. Pure Appl. Math.* 23 (1889), 376–8 = *Math. Papers*, Vol. 13, 26–8.
8. E. Coupy, Solution d'un problème appartenant à la géométrie de situation par Euler, *Nouv. Ann. Math*. 10 (1851), 106–19.
9. A. Crum Brown, On the theory of isomeric compounds, *Trans. Roy. Soc. Edinburgh* 23 (1964), 707–19.
10. A. De Morgan, Review of *The Philosophy of Discovery, Chapters Historical and Critical*, by W. Whewell, D.D., *Athenaeum* No. 1694 (1860), 501–3.
11. L. Euler, Solutio problematis ad geometriam situs pertinentis, *Comm. Acad. Sci. Imp. Petropol.* 8 (1736), 128–140 = *Opera Omnia (1)*, Vol. 7, 1–10.
12. L. Euler, Elementa doctrinae solidorum, *Novi Comm. Acad. Sci. Imp. Petropol*. 4 (1752–3), 109–40 = *Opera Omnia (1)*, Vol. 26, 72–93
13. L. Euler, Solution d'une question curieuse qui ne paroit soumise à aucune analyse, *Mém. Acad. Sci. Berlin* 15 (1759), 310–37 = *Opera Omnia (1)*, Vol. 7, 26–56.
14. E. Frankland, *Lecture Notes for Chemical Students*, London (1866).
15. W. R. Hamilton, Memorandum respecting a new system of roots of unity, *Philos. Mag. (4)* 12 (1856), 446.
16. P. J. Heawood, Map-colour theorem, *Quart. J. Pure Appl. Math*. 24 (1890), 332–8.
17. C. Hierholzer, Über die Möglichkeit, einen Lineanzug ohne Wiederholung und ohne Unterbrechnung zu umfahren, *Math. Ann*. 6 (1873), 30–2.
18. B. Hopkins and R. Wilson, The truth about Königsberg, *College Math. J*. 35 (2004), 198–207.
19. Major C. F. von Jaenisch, *Applications de l'Analyse Mathématiques au Jeu des Echecs*, 3 vols., St Petersburg (1862–3).
20. A. B. Kempe, On the geographical problem of the four colours, *Amer. J. Math*. 2 (1879), 193–200.
21. T. P. Kirkman, On the representation of polyedra, *Philos. Trans. Roy. Soc. London* 146 (1856), 413–18.
22. A.-M. Legendre, *Eléments de Géométrie* (1st edn), Firmin Didot, Paris (1794).
23. S.-A.-J. Lhuilier, Démonstration immédiate d'un théorème fondamental d'Euler sur les polyhèdres, et exceptions dont ce théorème est susceptible, *Mém. Acad. Imp. Sci. St. Pétersb*. 4 (1811), 271–301.

24. J. B. Listing, Vorstudien zur Topologie, *Göttinger Studien (Abt. 1) Math. Naturwiss. Abh.* 1 (1847), 811–75 = separately as book, Göttingen (1848).
25. J. B. Listing, Der Census räumliche Complexe, *Abh. K. Ges. Wiss. Göttingen Math. Cl.* 10 (1861–2), 97–182.
26. E. Lucas, *Récréations Mathématiques*, Vol. 1, Gauthier-Villars (1882).
27. A. C. Lunn and J. K. Senior, Isomerism and configuration, *J. Phys. Chem.* 33 (1929), 1027–79.
28. B. D. McKay, A note on the history of the four-colour conjecture, *J. Graph Theory* (2012), doi:10.1002/jgt.21662.
29. L. Poinsot, Sur les polygones et les polyèdres, *J. Ecole Polytech.* 4 (Cah. 10) (1809–10), 16–48.
30. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* 68 (1937), 145–254.
31. G. Pólya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer-Verlag (1987).
32. J.-C. Pont, *La Topologie Algébrique des Origines à Poincaré*, Bibliothèque de Philosophie Contemporaine, Presses Universitaires de France (1974).
33. H. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Arch. Math. Phys. (3)* 27 (1918), 142–4.
34. J. H. Redfield, The theory of group-reduced distributions, *Amer. J. Math.* 49 (1927), 433–55.
35. M. Reiss, Evaluation du nombre de combinaisons desquelles les 28 dés d'un jeu du domino sont susceptibles d'après la règle de ce jeu, *Ann. Mat. Pura. Appl. (2)* 5 (1871–3), 63–120.
36. W. W. Rouse Ball, *Mathematical Recreations and Problems of Past and Present Times* (later entitled *Mathematical Recreations and Essays*), Macmillan (1892).
37. H. Sachs, M. Stiebitz, and R. J. Wilson, An historical note: Euler's Königsberg letters, *J. Graph Theory* 12 (1988), 133–9.
38. J. J. Sylvester, Chemistry and algebra, *Nature* 17 (1877–8), 284 = *Math. Papers*, Vol. 3, 103–4.
39. J. J. Sylvester, On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, *Amer. J. Math.* 1 (1878), 64–125 = *Math. Papers*, Vol. 3, 148–206.
40. P. G. Tait, [Remarks on the colouring of maps], *Proc. Roy. Soc. Edinburgh* 10 (1878–80), 729.
41. W. T. Tutte, On Hamiltonian circuits, *J. London Math. Soc.* 21 (1946), 98–101.
42. A.-T. Vandermonde, Remarques sur les problèmes de situation, *Mém. Acad. Sci. (Paris)* (1771), 556–74.
43. O. Veblen, *Analysis Situs*, American Mathematical Society Colloquium Lectures 1916, New York (1922).
44. R. Wilson, *Four Colours Suffice*, Allan Lane (2002) (revised color edn., Princeton University Press, 2014).

Leonhard Euler (1707–83).

CHAPTER 9

Partitions

GEORGE E. ANDREWS

While Leibniz appears to have been the earliest to consider the partitioning of integers into sums, Euler was the first person to make truly deep discoveries. J. J. Sylvester was the next researcher to make major contributions, followed by Fabian Franklin. The 20th century saw an explosion of research with contributions from L. J. Rogers, G. H. Hardy, Percy MacMahon, Srinivasa Ramanujan, and Hans Rademacher.

Introduction

Gottfried Leibniz was apparently the first person to ask about partitions. In a letter of 1674 [45, p. 37] he asked Jacob Bernoulli about the number of 'divulsions' of integers; in modern terminology, he was asking the first question about partitions of integers. He observed that there are three partitions of 3 ($3, 2+1$, and $1+1+1$) as well as five of 4 ($4, 3+1, 2+2, 2+1+1$, and $1+1+1+1$). He then went on to observe that there are seven partitions of 5 and eleven of 6. This suggested that the number of partitions of any n might always be a prime; however, this *exemplum memorabile fallentis inductionis* is found out once one computes the fifteen partitions of 7. So even this first tentative exploration of partitions suggested a problem that is still open today:

Are there infinitely many integers n for which the total number of partitions of n is prime? (Put your money on 'yes'.)

From this small beginning we are led to a subject with many sides and many applications: the theory of partitions. The starting point is precisely that of Leibniz, put in modern notation.

Let $p(n)$ denote the number of ways in which n can be written as a sum of positive integers. A reordering of summands is not counted as a new partition, so $2+1+1$, $1+2+1$, and $1+1+2$ are considered the same partition of 4. As Leibniz noted,

$$p(3) = 3,\ p(4) = 5,\ p(5) = 7,\ p(6) = 11,\ p(7) = 15.$$

A number of questions can be asked about $p(n)$. How fast does it grow? What is its parity? Does it have special arithmetical properties? Are there efficient ways for computing $p(n)$? Is $p(n)$ prime infinitely often?

To give some order to an account of these questions, we organize the subject around the contributions of the great partition theorists: Euler, Sylvester, MacMahon, Rogers, Hardy, Ramanujan, and Rademacher. Each of these played a seminal role in the development of one or more themes in the history of partitions.

Euler and generating functions

In a letter from Philippe Naudé [27], Euler was asked to solve the problem of partitioning a given integer n into a given number of parts m. In particular, Naudé asked how many partitions there are of 50 into seven distinct parts.

The correct answer of 522 is not likely to be obtained by writing out all the ways of adding seven distinct positive integers to get 50. To solve this problem Euler used generating functions. Following Euler's lead, but using modern notation, we let $D(m, n)$ denote the number of partitions of n into m distinct parts. Then

$$\sum_{m,n\geq 0} D(m,n) z^m q^n = (1+zq^1)(1+zq^2)(1+zq^3)\cdots = \prod_{j=1}^{\infty}(1+zq^j)\,. \quad (1)$$

This identity becomes clear as we consider what happens when we multiply the terms on the right together. A typical term is

$$(zq^{i_1})(zq^{i_2})\cdots(zq^{i_j}) = z^j q^{i_1+i_2+\cdots+i_j},$$

which arises precisely from the partition with j distinct parts $i_1 + i_2 + \cdots + i_j$.

Noting that

$$\prod_{j=1}^{\infty}(1+zq^j) = (1+zq)\prod_{j=1}^{\infty}(1+(zq)q^j),$$

we can obtain a functional equation for the generating function of $D(m,n)$:

$$\sum_{m,n\geq 0} D(m,n)z^m q^n = (1+zq)\sum_{m,n\geq 0} D(m,n)z^m q^{n+m}.$$

Comparing the coefficients of $z^m q^n$ on both sides, we find that

$$D(m,n) = D(m,n-m) + D(m-1,n-m).$$

This equation allows an easy computation of the values of $D(m,n)$. Indeed, the following table is easily extended to include $D(7,50) = 522$:

$m \setminus n$	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0	1	1	1	1	1	1	1
2	0	0	0	1	1	2	2	3
3	0	0	0	0	0	0	1	1

Euler was naturally led from Naudé's question to an even more fundamental one. What is the generating function for $p(n)$, the total number of partitions of n? Here he applied the same principle that was so effective in computing $D(m,n)$ – namely,

$$\begin{aligned}\sum_{n=0}^{\infty} p(n)q^n &= (1+q^1+q^2+q^3+q^4+\cdots)\times(1+q^2+q^4+q^6+q^8+\cdots)\\ &\quad\times(1+q^3+q^6+q^9+q^{12}+\cdots)\times\cdots\\ &= \prod_{n=1}^{\infty}(1+q^n+q^{2n}+q^{3n}+q^{4n}+\cdots) = \prod_{n=1}^{\infty}\frac{1}{1-q^n}.\end{aligned}$$

At this point, Euler realized that a power series expansion for the product $\prod_{n=1}^{\infty}(1-q^n)$ would be essential for simplifying the computation of $p(n)$. He discovered empirically that

$$\begin{aligned}\prod_{n=1}^{\infty}(1-q^n) &= 1-q-q^2+q^5+q^7-q^{12}-q^{15}+q^{22}+q^{26}-\cdots \\ &= \sum_{n=-\infty}^{\infty}(-1)^n q^{n(3n-1)/2}\,.\end{aligned}$$

Some years after his empirical discovery, Euler managed to provide a proof of this himself; a modern exposition of his proof is given in [10]. This formula is now known as *Euler's pentagonal number theorem*. We shall examine Fabian Franklin's proof of it [29] in the next section.

Combining the pentagonal number theorem with the generating function for $p(n)$, one sees that

$$\left(\sum_{n=-\infty}^{\infty}(-1)^n q^{n(3n-1)/2}\right)\sum_{n=0}^{\infty}p(n)q^n = 1\,.$$

Comparing the coefficients of q^N on both sides of this last identity, Euler found the following recurrence for $p(N)$: $p(0)=1$, and

$$p(N) = p(N-1)+p(N-2)-p(N-5)-p(N-7)+\cdots\,, \text{ for } N>0.$$

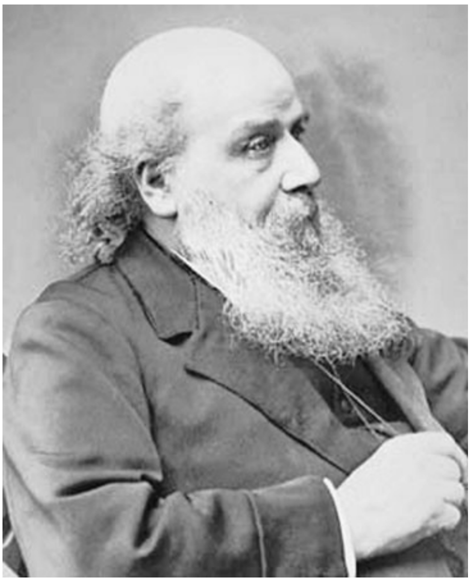
No one has ever found a more efficient algorithm for computing $p(N)$. It computes a full table of values of $p(n)$ for $n \le N$ in time $O(N^{3/2})$.

Euler's use of generating functions was the most important innovation in the entire history of partitions. Almost every discovery in partitions owes something to Euler's beginnings. Extensive accounts of the use of generating functions in the theory of partitions can be found in [5, Ch. 13], [9, Ch. 1, 2], [34, Ch. 7], [40, Ch. 19], and [54, Sect. 7, pp. 29–69]. In addition, a paper-by-paper summary of the entire history of partitions up to 1918 is found in Volume II of Dickson's *History of the Theory of Numbers* [25].

Sylvester and the intrinsic study of partitions

From the mid 18th century until the mid 19th century, little happened of significance in the study of partitions. Meanwhile, much was happening elsewhere in mathematics. Subjects such as the theory of complex variables and the theory of elliptic functions were born, and these would turn out to affect partitions profoundly. Great mathematicians such as Legendre, Gauss, Cauchy, and others would make new discoveries in their explications of Euler's work.

In the century between 1750 and 1850, the primary focus of research concerned explicit formulas for $p_k(n)$, the number of partitions of n into at most k parts. P. Paoli, A. De Morgan, J. F. W. Herschel, T. Kirkman, and H. Warburton studied $p_k(n)$ for small fixed values of k, and each produced a number of explicit formulas. We shall say more about these in a later section.



James Joseph Sylvester (1814–97).

James Joseph Sylvester was the next mathematician to provide truly new insight. In his magnum opus of 1884–6, 'A constructive theory of partitions, arranged in three acts, an interact, and an exodion' [68], Sylvester began with these words:

> In the new method of partitions it is essential to consider a partition as a definite thing, which end is attained by regularization of the succession of its parts according to some prescribed law. The simplest law for the purpose is that the arrangement of parts shall be according to their order of magnitude.

After considering several ways in which a partition may be given some sort of geometrical image, he asserted that the partition $5 + 5 + 4 + 3 + 3$, '... may be represented much more advantageously by the figure

$$\begin{array}{ccccc} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & \\ * & * & * & & \\ * & * & * & & \end{array},$$

This representation Sylvester called the *Ferrers graph* of the partition, after N. M. Ferrers, a Cambridge mathematician.

Sylvester noted immediately that one can count points in columns instead of rows; here this produces the partition $5 + 5 + 5 + 3 + 2$. The two partitions produced from such a graph are called *conjugates*. Thus, within the first two pages of his 83-page paper, Sylvester had inaugurated a brand new approach to partitions.

To appreciate the value of this new line of thought, we present Fabian Franklin's proof [29] of Euler's pentagonal number theorem. Franklin was one of Sylvester's students at Johns Hopkins University, and his proof illustrates the power of Sylvester's idea.

We begin by noting (as did Legendre) that the pentagonal number theorem can be reformulated purely as an assertion about partitions. If we set $z = -1$ in equation (1), we see that

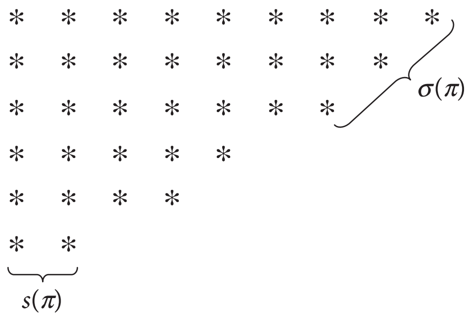
$$\prod_{j=1}^{\infty}(1 - q^j) = \sum_{m,n\geq 0} (-1)^m D(m, n)q^n.$$

Thus the coefficient of q^n is the difference between the number of partitions of n into an even number of distinct parts (say, $\Delta_e(n)$) and the number of partitions of n into an odd number of distinct parts (say, $\Delta_o(n)$). So Euler's pentagonal number theorem is equivalent to the assertion that

$$\Delta_e(n) - \Delta_o(n) = \begin{cases} (-1)^j, & \text{if } n = \frac{1}{2}j(3j \pm 1), \\ 0, & \text{otherwise.} \end{cases}$$

Franklin's idea for proving this was to find a one-to-one mapping between the partitions of n with an even number of parts (all distinct) and the partitions of n with an odd number of parts (all distinct). Of course, for the assertion to be valid, this mapping must run aground on an occasional exceptional case.

We proceed to examine partitions π with distinct parts. We define $s(\pi)$ to be the smallest summand in π, and $\sigma(\pi)$ to be the length of the longest sequence of consecutive integers appearing in π beginning with the largest part. For example, if π is $9+8+7+5+4+2$, then $s(\pi)=2$ and $\sigma(\pi)=3$ (corresponding to the sequence 9, 8, 7). We can provide images of $s(\pi)$ and $\sigma(\pi)$ when we look at the Ferrers graph of π:



Franklin then defined a transformation of partitions with distinct parts:

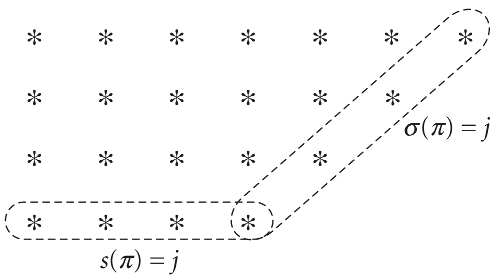
Case 1: $s(\pi) \le \sigma(\pi)$. In the Ferrers graph of π, move the points in $s(\pi)$ to the end of the first $s(\pi)$ rows. Thus the transformed partition arising when this move is applied to $9+8+7+5+4+2$ is $10+9+7+5+4$:

Case 2: $s(\pi) > \sigma(\pi)$. In the Ferrers graph of π, move the points in $\sigma(\pi)$ so that they form the smallest part of the transformed partition. Thus, if we consider the partition $10+9+7+5+4$, we see that $\sigma(\pi)=2$ and $s(\pi)=4$, and the transformed partition is $9+8+7+5+4+2$.

Note that Franklin's map changes the parity of the number of parts, and it appears not only to be one-to-one, but indeed to be an involution. If it really were both of these things, then we would have proved that the right-hand side of the formula for $\Delta_e(n) - \Delta_o(n)$ must always be 0.

However, Franklin's map gets into trouble whenever the two sets of points in the Ferrers graph defining $s(\pi)$ and $\sigma(\pi)$ are not disjoint. If they are disjoint, Franklin's map gives no problems: indeed, if $s(\pi) < \sigma(\pi)$ in Case 1 or $s(\pi) > \sigma(\pi) + 1$ in Case 2, everything still works.

Exceptional case 1: $s(\pi) = \sigma(\pi)$, and the defining sets for $s(\pi)$ and $\sigma(\pi)$ are not disjoint. For example,

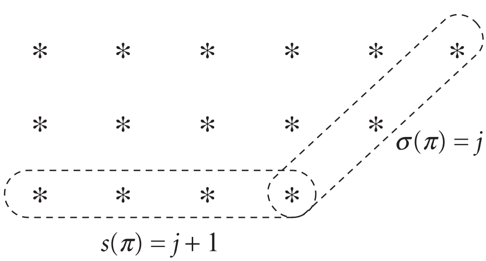


Clearly we cannot do the transformation required in Case 1. Thus, the partition with j distinct parts

$$j + (j + 1) + (j + 2) + \cdots + (2j - 1) = \tfrac{1}{2}j(3j - 1)$$

has no image under Franklin's mapping.

Exceptional case 2: $s(\pi) = \sigma(\pi) + 1$, and the defining sets for $s(\pi)$ and $\sigma(\pi)$ are not disjoint. For example,



Now we cannot do the transformation required in Case 2. Thus, the partition with j distinct parts

$$(j+1)+(j+2)+\cdots+2j=\tfrac{1}{2}j(3j+1)$$

has no image under Franklin's mapping.

We conclude that $\Delta_e(n)-\Delta_o(n)$ is 0, except when n is either $\frac{1}{2}j(3j-1)$ or $\frac{1}{2}j(3j+1)$, in which case it is $(-1)^j$. In other words, we have proved the formula for $\Delta_e(n)-\Delta_o(n)$, and consequently Euler's pentagonal number theorem.

Hans Rademacher termed Franklin's work the first major achievement of American mathematics. It may be viewed as the starting point for a variety of deep combinatorial studies of partitions.

Issai Schur's first proof [62, Sect. 3] of the (unknown to him) Rogers–Ramanujan identities owes much to Franklin; these identities form a major topic in our later section on Ramanujan. More recently, building on Schur's work, A. Garsia and S. Milne in 1981 produced a pure one-to-one correspondence proof of the Rogers–Ramanujan identities [30].

Perhaps the most striking achievement of this nature is the 1985 proof by D. Zeilberger and D. Bressoud [72] of the *q-Dyson conjecture* – namely, if

$$\prod_{1\le i<j\le n}\ \prod_{h=0}^{a_i-1}\left(1-\frac{x_i q^h}{x_j}\right)\left(1-\frac{x_j q^{h+1}}{x_i}\right)$$

is fully expanded, then the terms involving only powers of q but no x_is sum to the polynomial

$$\frac{(1-q)(1-q^2)(1-q^3)\cdots(1-q^{a_1+a_2+\cdots+a_n})}{\prod_{i=1}^{n}\{(1-q)(1-q^2)\cdots(1-q^{a_i})\}}.$$

Further accounts of these ideas and related work can be found in [6] and [11, Ch. 6 and 7].

Partitions representing other mathematical objects

It should not be surprising that partitions have a life outside their own intrinsic interest. After all, it is clear that whenever some set of n objects is grouped into subsets in which only the size of each subset is of significance, then the object of interest is a partition of n. For example, when Cauchy [21] first studied what in effect were the conjugacy classes of the symmetric group in 1845, their enumeration required classification by sizes of cycles. Thus, there are $p(n)$ conjugacy classes of the symmetric group S_n.

Classical invariant theory, as practised by Cayley [23], Sylvester [67], MacMahon [50, Vol. 2, Ch. 18] and others, bulges with the theory of partitions. For example, Stroh [65] solved a problem that had been considered extensively by Cayley, Sylvester, and MacMahon, when he showed that the generating function for perpetuants of a given degree θ (> 2) is

$$\frac{q^{2^{\theta-1}-1}}{(1-q^2)(1-q^3)\cdots(1-q^\theta)}.$$

Most important in this period was the realization that this external use of partitions was mutually beneficial. For example, from the work of Cauchy [20], Gauss [31], and others, it was determined that, for non-negative integers n and m, the expression

$$\begin{bmatrix} n+m \\ n \end{bmatrix} = \frac{(1-q^{n+1})(1-q^{n+2})\cdots(1-q^{n+m})}{(1-q)(1-q^2)\cdots(1-q^m)}$$

is actually a reciprocal polynomial in q of degree nm. Indeed, the coefficient of q^j is the number of partitions of j into at most m parts, each of which is less than or equal to n. These polynomials have come to be called *Gaussian polynomials* or *q-binomial coefficients*.

Empirically one finds that they are all unimodal; in fact, the coefficients of q^j form a non-decreasing sequence for $j \leq \frac{1}{2}nm$, and a non-increasing sequence for $j \geq \frac{1}{2}nm$. For example,

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 1 + q + 2q^2 + q^3 + q^4,$$

$$\begin{bmatrix} 7 \\ 3 \end{bmatrix} = 1 + q + 2q^2 + 3q^3 + 4q^4 + 4q^5 + 5q^6 + 4q^7 + 4q^8 + 3q^9 + 2q^{10} + q^{11} + q^{12}.$$

While this is an appealing oddity, viewed merely as a fact about partitions, it was of great importance in invariant theory – witness Sylvester's unbounded enthusiasm as he introduced his proof [66]:

I am about to demonstrate a theorem which has been waiting proof for the last quarter of a century and upwards. It is the more necessary that this should be done, because the theory has been supposed to lead to false conclusions, and its correctness has consequently been impugned. But, of the two suppositions that might be made to account for the observed discrepancy between the supposed consequences of the theorem and ascertained facts – one that the theory is false and the reasoning applied to it correct, the other that the theorem is true but that an error was committed in drawing certain deductions from it (to which one might add a third of the theorem and the reasoning upon it being both erroneous) – the wrong alternative was chosen. An error was committed in reasoning out certain supposed consequences of the theorem; but the theorem itself is perfectly true, as I shall show by an argument so irrefragable that it must be considered for ever hereafter safe from all doubt or cavil. It lies as the basis of the investigations begun by Professor Cayley in his *Second Memoir on Quantics*, which it has fallen to my lot, with no small labour and contention of mind, to lead to a happy issue, and thereby to advance the standards of the Science of Algebraical Forms to the most advanced point that has hitherto been reached. The stone that was rejected by the builders has become the chief corner-stone of the building.

The Gaussian polynomials were also observed by Dickson [26, p. 49] to count the total number of vector spaces of dimension n over a finite field of q elements (where q is now a prime power). Dickson proved this directly, but Knuth [43] was the first to point out exactly how the underlying partitions fit in.

To do this, Knuth considered all possible canonical bases for the k-dimensional subspaces of V_n, the vector space of dimension n over $GF(q)$, the Galois field of order q. Denoting the elements of V_n by $(x_1, x_2, \ldots, x_n)$, for $x_i \in GF(q)$, we see that a canonical basis for some subspace U consists of m vectors of the form

$$u_i = (u_{i1}, u_{i2}, \ldots, u_{in}), \text{ for } 1 \leq i \leq k,$$

satisfying

$$u_{in_i} = 1\,,\;\; u_{ij} = 0 \;\text{ for } j > n_i, \;\text{ and }\; u_{ln_i} = 0 \;\text{ for } l < i \text{ and } 1 \leq i \leq k,$$

where $n \geq n_1 > n_2 > \cdots > n_k \geq 1$. He then gave an example that revealed everything. If $n = 9$, $k = 4$, $n_1 = 8$, $n_2 = 5$, $n_3 = 3$, and $n_4 = 2$, then the canonical basis of U is $\{u_1, u_2, u_3, u_4\}$, where

$$\begin{aligned} u_1 &= (u_{11}, 0, 0, u_{14}, 0, u_{16}, u_{17}, 1, 0)\,, \\ u_2 &= (u_{21}, 0, 0, u_{24}, 1, 0, 0, 0, 0)\,, \\ u_3 &= (u_{31}, 0, 1, 0, 0, 0, 0, 0, 0)\,, \\ u_4 &= (u_{41}, 1, 0, 0, 0, 0, 0, 0, 0)\,. \end{aligned}$$

The pattern of the u_{ij} – namely,

$$\begin{matrix} * & * & * & * \\ * & * & & \\ * & & & \\ * & & & \end{matrix}$$

– is the Ferrers graph of the partition $4 + 2 + 1 + 1$. So each u_{ij} appearing above may be filled in $q^{4+2+1+1} = q^8$ ways. By this means we see clearly the correspondence with each partition with at most k parts being at most $n - k$, and Knuth's observation is clear.

Near the turn of the 19th century, the Revd Alfred Young, in a series of papers [71] on invariant theory, introduced partitions and variations thereof (now called *Young tableaux*) in what would come to be called the representation theory of the symmetric group (see [64]).

In the second half of the 20th century, applications mushroomed. J. W. B. Hughes [42] developed applications (in both directions) between Lie algebras and partitions. J. Lepowsky and R. L. Wilson [46] showed how to interpret and prove the Rogers–Ramanujan identities in Lie algebras, and a number of applications of partitions have arisen in physics (see [16]). Perhaps the most satisfying (and most surprising) was Rodney Baxter's solution of the hard hexagon model [14]; in simple terms, this work says that the Rogers–Ramanujan identities are crucial in studying the behaviour of liquid helium on a graphite plate [15]. It is fitting to close this section by noting that K. O'Hara [52] has discovered a purely partition-theoretic proof of Sylvester's theorem that the Gaussian polynomials are unimodal.

Asymptotics

One of the greatest surprises in the history of partitions was the formula of Hardy and Ramanujan [39] and Rademacher [55] for $p(n)$. However, it was not the first formula found for partition functions.

Cayley [22] and Sylvester [67] (anticipated by J. Herschel [41]) gave a number of formulas for $p_k(n)$ for small k, where $p_k(n)$ is the number of partitions of n into at most k parts (or, by conjugation, the number of partitions of n into parts each of which is at most k). For example,

$$p_2(n) = \left\lfloor \tfrac{1}{2}(n+1) \right\rfloor \quad \text{and} \quad p_3(n) = \left\{ \tfrac{1}{12}(n+3)^2 \right\},$$

where $\lfloor x \rfloor$ is the largest integer not exceeding x, and $\{x\}$ is the nearest integer to x.

Such results are fairly easy to prove, using nothing more powerful than the binomial series. For example,

$$\begin{aligned}\sum_{n=0}^{\infty} p_3(n)q^n &= \frac{1}{(1-q)(1-q^2)(1-q^3)} \\ &= \tfrac{1}{6}(1-q)^{-3} + \tfrac{1}{4}(1-q)^{-2} + \tfrac{1}{4}(1-q^2)^{-1} + \tfrac{1}{3}(1-q^3)^{-1} \\ &= \tfrac{1}{12}\sum_{n=0}^{\infty}(n+2)(n+1)q^n + \tfrac{1}{4}\sum_{n=0}^{\infty}(n+1)q^n + \tfrac{1}{4}\sum_{n=0}^{\infty} q^{2n} + \tfrac{1}{3}\sum_{n=0}^{\infty} q^{3n} \\ &= \sum_{n=0}^{\infty}\left(\tfrac{1}{12}(n+3)^2 - \tfrac{1}{3}\right)q^n + \tfrac{1}{4}\sum_{n=0}^{\infty} q^{2n} + \tfrac{1}{3}\sum_{n=0}^{\infty} q^{3n}.\end{aligned}$$

So $p_3(n)$ must be an integer that is in absolute value within $\frac{1}{3}$ of $\frac{1}{12}(n+3)^2$. Thus, the above formula for $p_3(n)$ is valid. An extensive account of such results and their history is given in [35].

In perhaps their most important joint paper [39], Hardy and Ramanujan found an asymptotic series for $p(n)$. The simplest special case of their result is the assertion that, as $n \to \infty$,

$$p(n) \sim \frac{1}{4n\sqrt{3}}\, e^{\pi\sqrt{2n/3}},$$

a result found independently by Uspensky [69] a few years later.

Statement of the main theorem.

Theorem. *Suppose that*

(1·71) $$\phi_q(n) = \frac{\sqrt{q}}{2\pi\sqrt{2}} \frac{d}{dn}\left(\frac{e^{C\lambda_n/q}}{\lambda_n}\right),$$

where C *and* λ_n *are defined by the equations* (1·53), *for all positive integral values of* q; *that* p *is a positive integer less than and prime to* q; *that* $\omega_{p,q}$ *is a 24q-th root of unity, defined when* p *is odd by the formula*

(1·721)

$$\omega_{p,q} = \left(\frac{-q}{p}\right) \exp\left[-\left\{\tfrac{1}{4}(2 - pq - p) + \tfrac{1}{12}\left(q - \frac{1}{q}\right)(2p - p' + p^2p')\right\}\pi i\right],$$

and when q *is odd by the formula*

(1·722).

$$\omega_{p,q} = \left(\frac{-p}{q}\right) \exp\left[-\left\{\tfrac{1}{4}(q - 1) + \tfrac{1}{12}\left(q - \frac{1}{q}\right)(2p - p' + p^2p')\right\}\pi i\right],$$

where (a/b) *is the symbol of Legendre and Jacobi*†, *and* p' *is any positive integer such that* $1 + pp'$ *is divisible by* q; *that*

(1·73) $$A_q(n) = \sum_{(p)} \omega_{p,q} e^{-2np\pi i/q};$$

and that α *is any positive constant, and* ν *the integral part of* $\alpha\sqrt{n}$.

Then

(1·74) $$p(n) = \sum_1^{\nu} A_q\phi_q + O(n^{-\frac{1}{4}}),$$

so that $p(n)$ *is, for all sufficiently large values of* n, *the integer nearest to*

(1·75) $$\sum_1^{\nu} A_q\phi_q.$$

Hardy and Ramanujan's exact formula for $p(n)$.

In 1937 Rademacher [55] improved the formula of Hardy and Ramanujan so that a convergent infinite series was found for $p(n)$ – namely,

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n) k^{1/2} \left[\frac{d}{dx}\left(\frac{\sinh\left((\pi/k)\left(\frac{2}{3}\left(x - \frac{1}{24}\right)\right)^{1/2}\right)}{\left(x - \frac{1}{24}\right)^{1/2}}\right)\right]_{x=n},$$

where

$$A_k(n) = \sum_{\substack{h \bmod k \\ (h,k)=1}} e^{-2\pi inh/k + i\pi s(h,k)}$$

with

$$s(h,k) = \sum_{\mu=1}^{k-1} \left((\mu/k) - \lfloor \mu/k \rfloor - \tfrac{1}{2}\right)\left((h\mu/k) - \lfloor h\mu/k \rfloor - \tfrac{1}{2}\right).$$

Now, this is not one of those mathematical formulas that elicits the response 'Just as I expected!' Indeed, it is really astounding: the proof consists of sheer wizardry. The contributions of both Hardy and Ramanujan were summarized as follows by J. E. Littlewood [48]:

> We owe the theorem to a singularly happy collaboration of two men, of quite unlike gifts, in which each contributed the best, most characteristic, and most fortunate work that was in him. Ramanujan's genius did have this one opportunity worthy of it.

Godfrey Harold Hardy (1877–1947) and Srinivasa Ramanujan (1887–1920).

Rademacher [55] also proved that if the series in the above equation for $p(n)$ is truncated after N terms, then the error is bounded in absolute value by

$$\frac{2\pi^2}{9\sqrt{3}}\, e^{\pi\sqrt{2n/3}\,/(N+1)}\, \frac{1}{N^{1/2}},$$

which tends to 0 as $N \to \infty$. This bound is quite crude; twenty-six terms of the series are required to get within one unit of the correct value of $p(200) =$ 3 972 999 029 388, a value calculated by hand by MacMahon. Actually, the result is within 0.004 of the correct answer after only eight terms are evaluated (see [59, p. 284]).

TABLE IV*: $p(n)$.

1...	1	51...	239943	101...	214481126	151...	45060624582
2...	2	52...	281589	102...	241265379	152...	49686288421
3...	3	53...	329931	103...	271248950	153...	54770336324
4...	5	54...	386155	104...	304801365	154...	60356673280
5...	7	55...	451276	105...	342325709	155...	66493182097
6...	11	56...	526823	106...	384276336	156...	73232243759
7...	15	57...	614154	107...	431149389	157...	80630964769
8...	22	58...	715220	108...	483502844	158...	88751778802
9...	30	59...	831820	109...	541946240	159...	97662728555
10...	42	60...	966467	110...	607163746	160...	107438159466
11...	56	61...	1121505	111...	679903203	161...	118159068427
12...	77	62...	1300156	112...	761002156	162...	129913904637
13...	101	63...	1505499	113...	851376628	163...	142798995930
14...	135	64...	1741630	114...	952050665	164...	156919475295
15...	176	65...	2012558	115...	1064144451	165...	172389800255
16...	231	66...	2323520	116...	1188908248	166...	189334822579
17...	297	67...	2679689	117...	1327710076	167...	207890420102
18...	385	68...	3087735	118...	1482074143	168...	228204732751
19...	490	69...	3554345	119...	1653668665	169...	250438925115
20...	627	70...	4087968	120...	1844349560	170...	274768617130
21...	792	71...	4697205	121...	2056148051	171...	301384802048
22...	1002	72...	5392783	122...	2291320912	172...	330495499613
23...	1255	73...	6185689	123...	2552338241	173...	362326859895
24...	1575	74...	7089500	124...	2841940500	174...	397125074750
25...	1958	75...	8118264	125...	3163127352	175...	435157697830
26...	2436	76...	9289091	126...	3519222692	176...	476715857290
27...	3010	77...	10619863	127...	3913864295	177...	522115831195
28...	3718	78...	12132164	128...	4351078600	178...	571701605655
29...	4565	79...	13848650	129...	4835271870	179...	625846753120
30...	5604	80...	15796476	130...	5371315400	180...	684957390936
31...	6842	81...	18004327	131...	5964539504	181...	749474411781
32...	8349	82...	20506255	132...	6620830889	182...	819876908323
33...	10143	83...	23338469	133...	7346629512	183...	896684817527
34...	12310	84...	26543660	134...	8149040695	184...	980462880430
35...	14883	85...	30167357	135...	9035836076	185...	1071823774337
36...	17977	86...	34262962	136...	10015581680	186...	1171432692373
37...	21637	87...	38887673	137...	11097645016	187...	1280011042268
38...	26015	88...	44108109	138...	12292341831	188...	1398341745571
39...	31185	89...	49995925	139...	13610949895	189...	1527273599625
40...	37338	90...	56634173	140...	15065878135	190...	1667727404093
41...	44583	91...	64112359	141...	16670689208	191...	1820701100652
42...	53174	92...	72533807	142...	18440293320	192...	1987276856363
43...	63261	93...	82010177	143...	20390982757	193...	2168627105469
44...	75175	94...	92669720	144...	22540654445	194...	2366022741845
45...	89134	95...	104651419	145...	24908858009	195...	2580840212973
46...	105558	96...	118114304	146...	27517052599	196...	2814570987591
47...	124754	97...	133230930	147...	30388671978	197...	3068829878530
48...	147273	98...	150198136	148...	33549419497	198...	3345365983698
49...	173525	99...	169229875	149...	37027355200	199...	3646072432125
50...	204226	100...	190569292	150...	40853235313	200...	3972999029388

Percy MacMahon's table of partition numbers, up to $p(200)$.

It would be hard to overstate the impact of this result on number theory in the 20th century. After Ramanujan's death in 1920, Hardy and Littlewood published a series of papers (see [38]) with the title *Some problems of Partitio Numerorum* (originally used by Euler for Chapter 16 of his *Introductio in Analysin Infinitorum* [28]). These papers gave deep results on sums of squares, Waring's problem, twin primes, the Goldbach conjecture, etc. After Rademacher made his contribution to $p(n)$, he and many of his school used his refinements and alternative work using Ford circles (see [56]) to prove a myriad of asymptotic theorems for the coefficients of modular forms and related functions.

The Rogers–Ramanujan identities

The history of partitions is filled with starts and stops. Euler's penetrating study of partitions stood for more than a hundred years before others made significant advances. One of the strangest stories surrounds two easily understood partition theorems: the *Rogers–Ramanujan identities*.

The partitions of n *into summands that differ from each other by at least* 2 *are equinumerous with the partitions into parts of the forms* $5m+1$ *and* $5m+4$.
The partitions of n *into summands each larger than* 1 *that differ from each other by at least* 2 *are equinumerous with the partitions into parts of the forms* $5m+2$ *and* $5m+3$.

We may express these results first as identities of the related generating functions:

$$1+\sum_{n=1}^{\infty}\frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)}=\prod_{n=0}^{\infty}\frac{1}{(1-q^{5n+1})(1-q^{5n+4})}$$

and

$$1+\sum_{n=1}^{\infty}\frac{q^{n^2+n}}{(1-q)(1-q^2)\cdots(1-q^n)}=\prod_{n=0}^{\infty}\frac{1}{(1-q^{5n+2})(1-q^{5n+3})}.$$

These latter two identities appeared in 1894 in a paper by L. J. Rogers that appeared in the *Proceedings of the London Mathematical Society*, hardly an obscure journal. The paper was entitled 'Second memoir on the expansion of certain infinite products' [61]. Apparently the mathematical public had lost interest somewhere in the first memoir on the *Expansion of Some Infinite Products* [60]. In any event, the paper was quite forgotten when nineteen years later an unknown Indian clerk, Srinivasa Ramanujan, sent these identities to G. H. Hardy.

At first glance these formulas look very much like ones that Euler had found, connected with the products that we described earlier. However, Hardy found that he was completely unable to prove them. He communicated them to Littlewood, MacMahon, and Perron, and no one could prove them. None of them thought of sending them to Rogers. MacMahon was the person who saw the series as generating functions for the two partition theorems, and he stated that (see [49, Vol. 2, p. 33]):

> This most remarkable theorem has been verified as far as the coefficient of q^{89} by actual expansion so that there is practically no reason to doubt its truth; but it has not yet been established.

Let us now turn to Hardy's account of the moment of illumination [37, p. 91]:

> The mystery was solved, trebly, in 1917. In that year Ramanujan, looking through old volumes of the *Proceedings of the London Mathematical Society*, came accidentally across Rogers's paper. I can remember very well his surprise, and the admiration which he expressed for Rogers's work. A correspondence followed in the course of which Rogers was led to a considerable simplification of his original proof. About the same time I. Schur, who was then cut off from England by the war, rediscovered the identities again. Schur published two proofs, one of which is "combinatorial" and quite unlike any other proof known.

Hardy wrote these words in 1940. For a while it seemed that these results were isolated curiosities. Fourteen years later, basing his remarks on the work of Lehmer [44] and Alder [1], Rademacher [57, p. 73] was to say 'It can be shown that there can be no corresponding identities for moduli higher than 5'. This turned out to be false. In 1961, B. Gordon [32] made the first step towards a full exploration of results of this type, by proving the following theorem.

Gordon's theorem. Let $A_{k,a}(n)$ *denote the number of partitions of* n *into parts not congruent to* 0, *or* $\pm a \pmod{2k+1}$. *Let* $B_{k,a}(n)$ *denote the number of partitions of* n *into parts of the form* $b_1 + b_2 + \cdots + b_j$, *where* $b_i \geq b_{i+1}$ *and* $b_i - b_{i+k-1} \geq 2$ *and at most* $a-1$ *of the* b_i *are* 1. *Then, for* $0 < a \leq k$ *and each* $n \geq 0$, $A_{k,a}(n) = B_{k,a}(n)$.

Gordon's theorem led to an explosion of results (see [2], [7], [8]). D. Bressoud made the first major combinatorial breakthroughs in the study of partition identities (see [18]). In recent years, K. Alladi and his collaborators [3] have found a substantial combinatorial theory of weighted words related to such results.

Ramanujan made many contributions to mathematics, which were carefully surveyed by Hardy in his book [37]. However, we should not fail to mention one more aspect of Ramanujan's discoveries; he discovered a number of divisibility properties of $p(n)$. Most notable are the following three [59, p. 210]:

$$p(5n+4) \equiv 0 \pmod 5,$$

$$p(7n+5) \equiv 0 \pmod 7,$$

$$p(11n+6) \equiv 0 \pmod{11}.$$

There are many other such results. Ramanujan conjectured infinite families of congruences, and these conjectures were proved (after the removal of some false ones) by Watson [70] and Atkin [13]. Most recently, K. Ono (in [53], among many other papers) has given us a picture of the depth and scope of these problems.

Other types of partitions

Earlier we alluded to some variations of partitions. The first mild variation is composition. A *composition* is a partition wherein different orders of summands count as different compositions. Thus, there are eight compositions of 4 – namely,

$$4,\ 3+1,\ 1+3,\ 2+2,\ 2+1+1,\ 1+2+1,\ 1+1+2,\ 1+1+1+1.$$

In 1876, Cayley [24] proved the first non-trivial theorem on compositions:

Cayley's theorem. Let F_n be the nth Fibonacci number. Then the number of compositions of n not using any 1*s is F_{n-1}.*

Major Percy Alexander MacMahon (1854–1929).

MacMahon gave compositions their name and scrutinized them thoroughly in [50, Vol. 1, Ch. 5]. As he developed his study of partitions, we can observe him almost blindly stumbling onto another form of partition, the *plane partitions* (see [50, pp. 1075–80]). Up to now, partitions have been linear or single-fold sums of integers: $n = \sum_{i=1}^{j} a_i$, where $a_i \geq a_{i+1}$. However, MacMahon found many intriguing properties associated with plane (or two-dimensional) partitions:

$$n = \sum_{i,j\geq 1} a_{ij}\,, \text{ where } a_{i,j} \geq a_{i,j+1} \text{ and } a_{i,j} \geq a_{i+1,j}.$$

Usually plane partitions are most easily understood when pictured as an array. For example, the six plane partitions of 3 are

$$\begin{array}{llllll} 3, & 21, & 2, & 111, & 11, & 1 \\ & & 1 & & 1 & 1 \\ & & & & & 1. \end{array}$$

Much to his surprise, MacMahon discovered (see [48, Vol. 1, p. 1071]), and proved seventeen years later (see [48, Vol. 1, Ch. 12]), that

$$\sum_{n=0}^{\infty} M(n)q^n = \prod_{n=1}^{\infty} \frac{1}{(1-q^n)^n}.$$

At first MacMahon believed that comparably interesting discoveries awaited higher-dimensional partitions. However, he later noted that such hopes were in vain (see [50, Vol. 1, p. 1168]).

Many subsequent discoveries have been made for plane partitions. A beautiful development of recent work has been given by Richard Stanley, in two papers [63] and an excellent book [64].

Further leads to the history

A short chapter like this must slight much of the history of partitions. Many favourite topics have received little or no attention. So in this final section we mention some historical sources where one may find a more detailed treatment of various aspects of the history of partitions.

First and foremost is Chapter III of Volume 2 of L. E. Dickson's *History of the Theory of Numbers* [25], which cites every paper on partitions known up to 1916. H. Ostmann's *Additive Zahlentheorie*, Volume 1, Chapter 7 [54] contains a fairly full account of progress in the first half of the 20th century. Reviews of all papers on partitions from 1940 to 1983 can be found in [47] and [36], in Chapter P. Inasmuch as MacMahon was a seminal and lasting influence in partitions, one should examine Volume 1 of his *Collected Papers* [50, Vol. 1]; partitions were used in his *Combinatory Analysis* [49], and each chapter is introduced with some history and a bibliography of work since his death in 1930.

Besides these major works, there have been a number of survey articles with extensive histories. H. Gupta provided a general survey in [33]. Partition identities are handled in [2], [4], [7], and [12]. Richard Stanley gave a history of plane partitions in [63]. Applications in physics are discussed by Berkovich and McCoy in [16] (see also [11, Ch. 8]).

Finally, there are books that have some of the history of partitions. Andrews [9] is devoted entirely to partitions, and the Notes sections concluding each chapter have extensive historical references. Bressoud [19] has recently published the history of the alternating sign matrix conjecture, an appealing and well-told tale that is tightly bound up with the theory of partitions. Ramanujan's amazing contributions to partitions (as well as many other aspects of number theory) have been chronicled by G. H. Hardy [37], and most thoroughly by

B. Berndt in five volumes [17]. Books with chapters on partitions include those of Gupta [34, Ch. 7–10], Hardy and Wright [40, Ch. 19], Macdonald [51, Ch. 1, Sect. 1], Rademacher [57, Parts I and III], Rademacher [58, Ch. 12–14], and Stanley [64, Ch. 7].

References

1. H. L. Alder, The nonexistence of certain identities in the theory of partitions and compositions, *Bull. Amer. Math. Soc.* 54 (1948), 712–22.
2. H. L. Alder, Partition identities – from Euler to the present, *Amer. Math. Monthly* 76 (1969), 733–46.
3. K. Alladi, Refinements of Rogers–Ramanujan type identities, *Special Functions, q-Series and Related Topics*, *Fields Inst. Commun.* 14 (1995), 1–35.
4. G. E. Andrews, A polynomial identity which implies the Rogers–Ramanujan identities, *Scripta Math.* 28 (1970), 297–305.
5. G. E. Andrews, *Number Theory*, W. B. Saunders (1971) (reprinted by Dover, 1994).
6. G. E. Andrews, Two theorems of Gauss and allied identities proved arithmetically, *Pacific J. Math.* 41 (1972), 563–78.
7. G. E. Andrews, Partition identities, *Adv. Math.* 9 (1972), 10–51.
8. G. E. Andrews, On the general Rogers–Ramanujan theorem, *Mem. Amer. Math. Soc.* 152 (1974), 86 pp.
9. G. E. Andrews, The theory of partitions, *Encyclopedia of Mathematics and its Applications*, Vol. 2, Addison-Wesley (1976) (reissued by Cambridge University Press, 1985, 1998).
10. G. E. Andrews, Euler's pentagonal number theorem, *Math. Mag.* 56 (1983), 279–84.
11. G. E. Andrews, *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra*, C.B.M.S. Regional Conference Lecture Series 66, American Mathematical Society (1986).
12. G. E. Andrews, *On the proofs of the Rogers–Ramanujan identities, from q-series and partitions* (ed. D. Stanton), I.M.A. Volumes in Mathematics 18, Springer (1989), 1–14.
13. A. O. L. Atkin, Proof of a conjecture of Ramanujan, *Glasgow Math. J.* 8 (1967), 14–32.
14. R. J. Baxter, Hard hexagons: exact solution, *J. Phys. A* 13 (1980), L61–L70.
15. R. J. Baxter, Rogers–Ramanujan identities in the hard hexagon model, *J. Stat. Phys.* 26 (1981), 427–52.
16. A. Berkovich and B. M. McCoy, Rogers–Ramanujan identities: a century of progress from mathematics to physics, *Documenta Math.*, Extra Vol. III, International Congress of Mathematicians, Berlin (1998), 163–72.
17. B. Berndt, *Ramanujan's Notebooks, Parts I–V*, Springer (1985, 1989, 1991, 1994, 1998).

18. D. M. Bressoud, Analytic and combinatorial generalizations of the Rogers–Ramanujan identities, *Mem. Amer. Math. Soc.* 227 (1980).
19. D. M. Bressoud, *Proofs and Confirmations*, Cambridge University Press (1999).
20. A. Cauchy, Mémoire sur les functions dont plusieurs valeurs sont liées entre elles par une équation linéaire, et sur diverses transformations de produits composées d'un nombre indéfini de facteurs, *Oeuvres Complètes*, Ser. 1, Vol. 8 (1893), 46.
21. A. Cauchy, Mémoire sur la résolution des équations linéaires symboliques, et sur les conséquences remarquables que cette résolution entraîne après elle dans la theorie des permutations, *Comptes Rendus* 21 (1845), 1123 = *Oeuvres Complètes*, Ser. 1, Vol. 9 (1896), 417–30.
22. A. Cayley, Researches on the partition of numbers, *Philos. Trans. Roy. Soc.* 145 (1855), 127–40 = *Math. Papers*, Vol. 2, 235–49.
23. A. Cayley, A second memoir on quantics, *Philos. Trans. Roy. Soc.* 146 (1856), 101–26 = *Math. Papers*, Vol. 2, 250–75.
24. A. Cayley, Theorems in trigonometry and on partitions, *Messenger of Math.* 5 (1876) = *Math. Papers*, Vol. 10, 16.
25. L. E. Dickson, *History of the Theory of Numbers*, Vol. 2, Carnegie Institution, Washington, DC (1920) (reprinted by Chelsea, 1952).
26. L. E. Dickson, *Linear Groups* (reprinted by Dover, 1958).
27. L. Euler, Observationes analyticae variae de combinationibus, *Comm. Acad. Imp. Petropol.* 13 (1741–3, 1751), 64–93.
28. L. Euler, *Introductio in Analysin Infinitorum*, Marcun-Michaelem Bousquet, Lausanne (1748) (reprinted, *Culture et Civilisation*, Brussels, 1967).
29. F. Franklin, Sur le développement du produit infini $(1-x)(1-x^2)(1-x^3)\dots$, *Comptes Rendus* 82 (1881), 448–50.
30. A. Garsia and S. Milne, A Rogers–Ramanujan bijection, *J. Combin. Theory (A)* 31 (1981), 289–339.
31. C. F. Gauss, *Quarundan Serierum Singularum, Werke*, Vol. 2, Königlichen Gesell.-der Wissen., Göttingen (1876) (originally published in 1811).
32. B. Gordon, A combinatorial generalization of the Rogers–Ramanujan identities, *Amer. J. Math.* 83 (1961), 393–99.
33. H. Gupta, Partitions – a survey, *J. Res. Nat. Bur. Standards* 74B (1970), 1–29.
34. H. Gupta, *Selected Topics in Number Theory*, Abacus (1980).
35. H. Gupta, A. E. Gwyther, and J. C. P. Miller, Tables of partitions, *Roy. Soc. Math. Tables* 4 (1958).
36. R. Guy, *Reviews in Number Theory, 1973–1983*, 6 vols, American Mathematical Society (1984).
37. G. H. Hardy, *Ramanujan*, Cambridge University Press (reprinted by Chelsea, 1999).
38. G. H. Hardy, *Collected Papers*, Vol. I (ed. J. E. Littlewood), Oxford University Press (1966).
39. G. H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, *Proc. London Math. Soc. (2)* 17 (1918), 75–115 = *Collected Papers of S. Ramanujan* (1927), 276–309 (reprinted by Chelsea, 1982).

40. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th edn, Oxford University Press (2008).
41. J. W. F. Herschel, On circulating functions and on the integration of a class of equations of finite differences into which they enter as coefficients, *Philos. Trans. Roy. Soc.* 108 (1818), 144–68.
42. J. W. B. Hughes, Lie algebraic proofs of some theorems on partitions, *Number Theory and Algebra*, Academic Press (1977), 135–55.
43. D. Knuth, Subspaces, subsets and partitions, *J. Combin. Theory (A)* 10 (1971), 178–80.
44. D. H. Lehmer, Two nonexistence theorems on partitions, *Bull. Amer. Math. Soc.* 52 (1946), 538–44.
45. G. W. Leibniz, *Math. Schriften*, Vol. IV, 2, Specimen de divulsionibus aequationum … Letter 3 dated 2 Sep. 1674 (see D. Mahnke, Leibniz auf der Suche nach einer allgemeinen Primzahlgleichung, *Bibliotheca Math*. 13 (1912–13), 29–61).
46. J. Lepowsky and R. L. Wilson, A Lie-theoretic interpretation and proof of the Rogers–Ramanujan identities, *Adv. Math.* 45 (1982), 21–72.
47. W. J. LeVeque, *Reviews in Number Theory, 1940–1972*, 6 vols, American Mathematical Society (1974).
48. J. E. Littlewood, *Littlewood's Miscellany* (ed. B. Bollobás), Cambridge University Press (1986).
49. P. A. MacMahon, *Combinatory Analysis*, 2 vols, Cambridge University Press (1916) (reprinted by Chelsea, 1960).
50. P. A. MacMahon, *Collected Papers*, 2 vols. (ed. G. E. Andrews), MIT Press (1978, 1986).
51. I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford University Press (1979).
52. K. O'Hara, Unimodality of the Gaussian coefficients: a constructive proof, *J. Combin. Theory (A)* 53 (1990), 29–52.
53. K. Ono, The partition function in arithmetic progressions, *Math. Ann*. 312 (1998), 251–60.
54. H. Ostmann, *Additive Zahlentheorie*, Vol. 1, Springer (1956).
55. H. Rademacher, On the partition function $p(n)$, *Proc. London Math. Soc. (2)* 43 (1937), 241–54.
56. H. Rademacher, On the expansion of the partition function in a series, *Ann. Math.* 44 (1943), 416–22.
57. H. Rademacher, *Lectures on Analytic Number Theory*, Tata Institute (1954).
58. H. Rademacher, *Topics in Analytic Number Theory*, Grundlehren der math. Wissenschaften 169, Springer (1973).
59. S. Ramanujan, *Collected Papers*, Cambridge University Press (1927) (reprinted by Chelsea, 1962).
60. L. J. Rogers, On the expansion of some infinite products, *Proc. London Math. Soc.* 24 (1893), 337–52.
61. L. J. Rogers, Second memoir on the expansion of certain infinite products, *Proc. London Math. Soc.* 25 (1894), 318–43.

62. I. Schur, Ein Beitrag zur additiven Zahlentheorie und zur Theorie der Kettenbrüche, *Sitzungsber. Preuss, Akad. Wiss. Phys.-Math. Klasse* (1917), 302–21 = *Ges. Abh.*, Vol. 2, Springer (1973), 117–36.
63. R. P. Stanley, Theory and application of plane partitions, I, II, *Stud. Appl. Math.* 50 (1971), 167–88, 259–79.
64. R. P. Stanley, *Enumerative Combinatorics*, Vol. 2, Cambridge University Press (1999).
65. E. Stroh, Ueber die Symbolische Darstellung den Grundsyzyganten einer binären Form sechster Ordnung und eine Erweckerung der Symbolik von Clebsch, *Math. Ann.* 36 (1890), 263–303.
66. J. J. Sylvester, Proof of the hitherto undemonstrated fundamental theorem of invariants, *Philos. Mag.* 5 (1878), 178–88 = *Collected Papers*, Vol. III (1909), 117–26.
67. J. J. Sylvester, On subinvariants, that is, semi-invariants to binary quantics of an unlimited order, *Amer. J. Math.* 5 (1882), 79–136 = *Collected Papers*, Vol. III (1909), 568–622.
68. J. J. Sylvester, A constructive theory of partitions, arranged in three acts, an interact and an exodion, *Amer. J. Math.* 5–6 (1884–6), 251–330, 334–36 = *Collected Papers*, Vol. IV (1912), 1–83.
69. J. V. Uspensky, Asymptotic formulae for numerical functions which occur in the theory of partitions [in Russian], *Bull. Acad. Sci. USSR (6)* 14 (1920), 199–218.
70. G. N. Watson, Ramanujan's Vermutung über Zerfällungsanzahlen, *J. reine angew. Math.* 179 (1938), 97–128.
71. A. Young, *Collected Papers* (ed. G. de B. Robinson), University of Toronto Press (1977).
72. D. Zeilberger and D. M. Bressoud, A proof of Andrews' q-Dyson conjecture, *Discrete Math.* 54 (1985), 201–24.

The Revd Thomas Penyngton Kirkman (1806–95).

CHAPTER 10

Block designs

NORMAN BIGGS AND ROBIN WILSON

This chapter outlines the origins of design theory, with particular reference to the 'Steiner triple systems', the pioneering work of Thomas Kirkman, and early contributions by a number of writers on the 'fifteen schoolgirls problem'. Connections are also made with finite projective planes and with the design of experiments [1].

Triple systems

In the *Lady's and Gentleman's Diary* for 1844 the editor, Wesley Woolhouse, posed the following problem:

> *Prize Quest. (1733); by the Editor.* Determine the number of combinations that can be made out of *n* symbols, *p* symbols in each; with this limitation, that no combination of *q* symbols, which may appear in any one of them shall be repeated in any other.

The *Diary* was

> designed principally for the amusement and instruction of students in mathematics: comprising many useful and entertaining particulars, interesting to all persons engaged in that delightful pursuit,

and its readers were invited to send in solutions of the questions posed. Prize Question 1733 is about arranging things: we are asked to arrange a number (n)

of symbols into groups of p elements in such a way that a particular condition is satisfied.

Unfortunately, Woolhouse's problem can be interpreted in various ways, and just two solutions were received. One misunderstood the problem altogether, while the other, by Septimus Tebay of the Gas Works in Preston, gave the answer $C(n, q) / C(p, q)$; if it is assumed that the arrangement is possible (which is not always the case), this is the correct number.

Because this was not the interpretation that Woolhouse had intended, in 1846 he duly presented his readers with a simplified challenge, corresponding to the specific case $p = 3, q = 2$:

Question 1760. How many triads can be made out of *n* symbols, so that no pair of symbols shall be comprised more than once amongst them?

Here is an example for $n = 7$, where the seven symbols are the numbers 1, 2, 3, 4, 5, 6, 7, and the triads (triples) are arranged vertically:

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

The condition to be satisfied is that no pair of numbers may occur together more than once. In this example, each pair appears exactly once – for example, the numbers 3 and 5 appear together in the second triple, while the numbers 2 and 6 appear together in the sixth triple.

Such arrangements of numbers are now usually called *Steiner triple systems*, for reasons that will be explained shortly. We shall denote a Steiner triple system with n symbols by $S(n)$. The example $S(7)$ given above is easy to construct, since from the first triple $(1, 2, 4)$ we obtain each successive triple by adding 1 to each number, always following 7 by 1; such systems are called *cyclic systems*.

Another example of a triple system, an $S(9)$, is shown below; it has nine symbols and twelve triples. Again, each pair of numbers appears in exactly one triple – for example, the numbers 3 and 8 appear together in the eighth triple:

1	1	1	1	2	2	2	3	3	3	4	7
2	4	5	6	4	5	6	4	5	6	5	8
3	7	9	8	9	8	7	8	7	9	6	9

For which values of n do such Steiner triple systems exist? Note that, in a triple system with n symbols, each number appears $f = \frac{1}{2}(n - 1)$ times, because the other $n-1$ numbers appear in pairs with it. Also, when there are t triples, the total number of entries in the system is $3t = nf$, so $t = \frac{1}{6}n(n-1)$. Since this is an integer, n must be one of the numbers in the sequence

$$7, 9, 13, 15, 19, 21, 25, 27, \ldots ;$$

these are the numbers of the form $6k+1$ and $6k+3$, where k is an integer. Initially it was not clear whether a triple system can be constructed for each such number n, but this is indeed the case: there is essentially one system for $n = 7$ and for $n = 9$, and there are two systems for $n = 13$, eighty for $n = 15$, and millions for all higher values ($n = 19, 21, 25, 27, \ldots$).

For future reference, we call a triple system *resolvable* if its triples can be rearranged into subsystems, each containing all n numbers. For example, the above system $S(9)$ can be rearranged into four sets of three triples, each set containing all nine numbers:

1	4	7	1	2	3	1	2	3	1	2	3
2	5	8	4	5	6	6	4	5	5	6	4
3	6	9	7	8	9	8	9	7	9	7	8

Such an arrangement is possible only when n is divisible by 3, which means that n must be of the form $6k+3$.

Triple systems first appeared in the work of the German geometer Julius Plücker. In his 1835 book on analytic geometry [38] he observed that 'a general plane cubic curve has 9 points of inflection that lie in threes on 12 lines'; moreover, 'given any two points of the system, exactly one of the lines passes through them both'; this is just a description of the system that we have called $S(9)$. Plücker presented the system explicitly, and in a footnote he observed that:

> If a system $S(n)$ of n points can be arranged in triples, so that any two points lie in just one triple, then n has the form $6k+3 \ldots$.

In a later book [39] he corrected his mistake, adding the possibility $n = 6k+1$.

It is a matter of speculation as to how Wesley Woolhouse became interested in triple systems, but one possibility is that James Joseph Sylvester, who had a life-long interest in combinatorial systems and wrote a paper [44] on 'combinatorial aggregation' in 1844, knew of Plücker's work and mentioned it to Woolhouse. We shall have more to say about Sylvester later.

Kirkman's 1847 paper

The Revd Thomas Penyngton Kirkman was a keen mathematician who had studied in Dublin, and was rector of the small parish of Croft-with-Southworth in Lancashire. His parochial duties took up little of his time, and he concentrated much effort on his mathematical researches, especially on algebraic and combinatorial topics, being elected to a Fellowship of the Royal Society in the process.

On 15 December 1846 Kirkman read a paper to the Literary and Philosophical Society of Manchester, entitled 'On a problem in combinations'. In this pioneering paper, published the following year in the *Cambridge and Dublin Mathematical Journal* [21], he showed how to construct a Steiner triple system $S(n)$ for each positive integer n of the form $6k + 1$ and $6k + 3$, a substantial achievement.

In order to do this, Kirkman introduced a supplementary system $D(2m)$, which is an arrangement into $2m - 1$ columns of the $C(2m, 2)$ pairs of $2m$ symbols; these systems had earlier been considered by Sylvester in his 1844 paper. Kirkman's system D_8 is shown below; it can also be regarded as a colouring of the edges of the complete graph K_8 with vertices h, i, k, l, m, n, o, p, where the columns list those sets of four edges that are assigned the same colour:

hi	*hk*	*hl*	*hm*	*hn*	*ho*	*hp*
kl	*il*	*ik*	*in*	*im*	*ip*	*io*
mn	*mo*	*mp*	*ko*	*kp*	*mk*	*nk*
op	*np*	*no*	*lp*	*lo*	*nl*	*ml*

He then gave two constructions, in which he used these systems to extend smaller Steiner systems to larger ones.

- From a Steiner triple system $S(n)$ and a supplementary system $D(n + 1)$ he constructed a Steiner triple system $S(2n + 1)$. For example, from the system $S(7)$ and the above design $D(8)$ he constructed a system $S(15)$.
- From a triple system $S(2n + 1)$ he removed two symbols and some triples to obtain a partial system $S^*(2n - 1)$ in which certain pairs do not appear; then, from a partial system $S^*(m + 1)$ he constructed a triple system $S(2m + 1)$. For example, from the system $S(7)$ he successively obtained $S^*(5)$ and $S(9)$.

By combining these constructions, Kirkman was able to construct a Steiner triple system with n symbols for every n of the form $6k + 1$ or $6k + 3$.

The contributions of Steiner

In 1853 the Swiss geometer Jakob Steiner wrote a short note [43] on triple systems, a topic that he had probably encountered while studying Plücker's work. In this note Steiner correctly observed that triple systems with n points can exist only when n has the form $6k + 1$ or $6k + 3$. He also asked whether triple systems could be constructed for all such numbers, unaware that Kirkman had completely solved this problem six years earlier. This lack of awareness may have arisen from the fact that the *Cambridge and Dublin Mathematical Journal*, though well known in Britain, was little known on the Continent. The situation was further complicated when M. Reiss [41] solved Steiner's problem using methods very similar to those of Kirkman, causing the latter to complain sarcastically [28]:

> ... how did the Cambridge and Dublin Mathematical Journal ... contrive to steal so much from a later paper in Crelle's Journal, Vol. LVI., p. 326, on exactly the same problem in combinations?

The term 'Steiner triple system' was coined much later by Ernst Witt [52]. Thus, not only did Kirkman fail to gain the credit for 'Hamiltonian' graphs, which should rightly have been named after him (see Chapter 8), but he also missed out on receiving the credit for his fundamental work on the construction of triple systems.

Kirkman's schoolgirls problem

While preparing his 1847 paper, Kirkman noticed that the thirty-five triples of his system $S(15)$ can be split into subsystems, each containing all fifteen points – that is, it is a resolvable triple system. In the *Lady's and Gentleman's Diary* for 1850 [22], intermingled with challenges on the sons of Noah and the origins of April Fool's Day, he proposed a recreational form of this observation, now known as *Kirkman's schoolgirls problem*:

> Fifteen young ladies in a school walk out three abreast for seven days in succession: it is required to arrange them daily, so that no two shall walk twice abreast.

If there had been only nine young ladies, we could have used the resolvable system $S(9)$ displayed earlier. The four subsystems correspond to four days: on the first day, 1 walks with 2 and 3, 4 walks with 5 and 6, and 7 walks with 8 and 9; on the second day, 1 walks with 4 and 7, and so on; and no two young ladies walk together more than once.

The problem of the fifteen young ladies also appeared in the *Educational Times* 'thus versified by a lady' [3]:

A governess of great renown
Young ladies had fifteen,
Who promenaded near the town,
Along the meadows green.

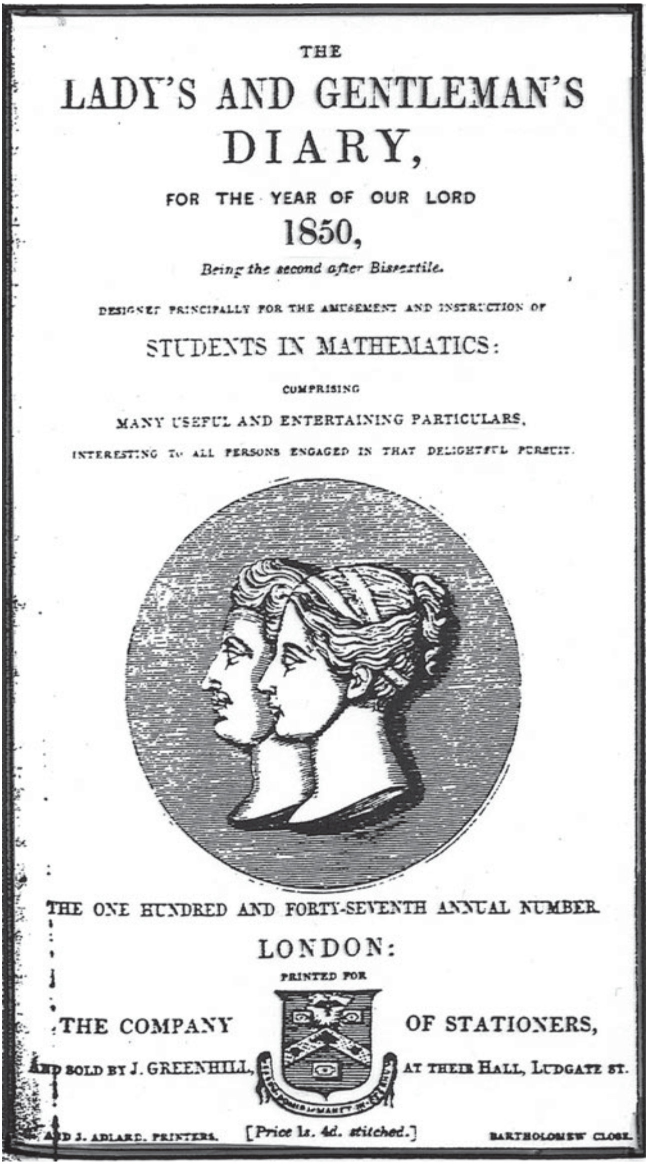
But as they walked
They tattled and talked,
In chosen ranks of three,
So fast and so loud,
That the governess vowed
It should no longer be.

So she changed them about,
For a week throughout,
In threes, in such a way
That never a pair
Should take the air
Abreast on a second day;
And how did the governess manage it, pray?

A solution of the fifteen schoolgirls problem, listing the five triples for each day, is shown below.

Monday:	1–2–3	4–5–6	7–8–9	10–11–12	13–14–15
Tuesday:	1–4–7	2–5–8	3–12–15	6–10–14	9–11–13
Wednesday:	1–10–13	2–11–14	3–6–9	4–8–12	5–7–15
Thursday:	1–5–11	2–6–12	3–7–13	4–9–14	8–10–15
Friday:	1–8–14	2–9–15	3–4–10	6–7–11	5–12–13
Saturday:	1–6–15	2–4–13	3–8–11	5–9–10	7–12–14
Sunday:	1–9–12	2–7–10	3–5–14	6–8–13	4–11–15

Notice that, as before, any two schoolgirls walk together exactly once; for example, schoolgirls 3 and 10 walk together on Friday.


THE

LADY'S AND GENTLEMAN'S

DIARY,

FOR THE YEAR OF OUR LORD

1850,

Being the second after Bissextile.

DESIGNED PRINCIPALLY FOR THE AMUSEMENT AND INSTRUCTION OF

STUDENTS IN MATHEMATICS:

COMPRISING

MANY USEFUL AND ENTERTAINING PARTICULARS,

INTERESTING TO ALL PERSONS ENGAGED IN THAT DELIGHTFUL PURSUIT.

THE ONE HUNDRED AND FORTY-SEVENTH ANNUAL NUMBER.

LONDON:

PRINTED FOR

THE COMPANY OF STATIONERS,

AND SOLD BY J. GREENHILL, AT THEIR HALL, LUDGATE ST.

AND J. ADLARD, PRINTERS, BARTHOLOMEW CLOSE.

[Price 1s. 4d. stitched.]

The *Lady's and Gentleman's Diary* for 1850.

The schoolgirls problem proved to be more successful than the 1844 Prize Question, and two solutions appeared in the *Diary* for 1851: one by Kirkman himself, and one supposedly obtained independently by Mr Bills of Newark, Mr Jones of Chester, Mr Wainman of Leeds, and Mr Levy of Hungerford; how they all produced exactly the same solution is unclear. Kirkman claimed that his own solution was 'the symmetrical and only possible solution', but he was wrong; another symmetrical solution, different from his, had been obtained a few months earlier by Arthur Cayley [11].

Kirkman himself described several variations of the problem [24]. He solved the corresponding problem for nine young ladies, believed it to be impossible for twenty-one young ladies, and asserted the following results:

> Sixteen young ladies can all walk out four abreast, till every three have once walked abreast; so can thirty-two, and so can sixty-four young ladies; so can 4^n young ladies.

In 1852 Spottiswoode [42] extended the problem to $2^{2n} - 1$ young ladies, walking in threes over $2^{2n-1} - 1$ days – that is, to $15, 63, 225, 1023, 4095, \ldots$ young ladies, where the numbers of days are $7, 31, 127, 511, 2047, \ldots$, respectively.

Also in 1852 a new type of solution was produced by the Revd Robert Anstice [4], who had studied mathematics in Oxford. Unlike the previous somewhat ad hoc attempts, he sought systems that involved some structure, and succeeded in finding a *cyclic* solution. He also found solutions for any $n = 2p + 1$, where p is a prime number of the form $6k + 1$.

Anstice's solution for the case $n = 15$ is given below, with the schoolgirls denoted by 0–6 in normal type, **0**–**6** in bold face, and the symbol ∞ (infinity). Notice that once we have the arrangement for Monday, we can obtain the arrangement for each successive day by adding 1, always following 6 by 0, and leaving ∞ unchanged.

Monday:	∞–0–**0**	**1**–2–**3**	1–**4**–**5**	3–5–6	4–**2**–**6**
Tuesday:	∞–1–**1**	**2**–3–**4**	2–**5**–**6**	4–6–0	5–**3**–**0**
Wednesday:	∞–2–**2**	**3**–4–**5**	3–**6**–**0**	5–0–1	6–**4**–**1**
Thursday:	∞–3–**3**	**4**–5–**6**	4–**0**–**1**	6–1–2	0–**5**–**2**
Friday:	∞–4–**4**	**5**–6–**0**	5–**1**–**2**	0–2–3	1–**6**–**3**
Saturday:	∞–5–**5**	**6**–0–**1**	6–**2**–**3**	1–3–4	2–**0**–**4**
Sunday:	∞–6–**6**	**0**–1–**2**	0–**3**–**4**	2–4–5	3–**1**–**5**

Further cyclic solutions were produced by the American mathematician Benjamin Peirce [37]. Indeed, he showed that there are just three types of cyclic solution, the ones obtained respectively by Anstice, Kirkman, and Cayley. These three types are determined by their Monday schedules, as follows.

Anstice:	∞–0–**0**	**1**–2–**3**	1–**4**–**5**	3–5–6	4–**2**–**6**
Kirkman:	∞–0–**0**	**1**–**3**–4	2–**4**–**5**	3–5–6	1–**2**–**6**
Cayley:	∞–0–**0**	**1**–5–**6**	**3**–**4**–6	1–2–4	3–**2**–**5**

The contributions of Sylvester

At this point we must refer again to the brilliant but eccentric James Joseph Sylvester, whose chequered career, on both sides of the Atlantic, culminated in his election as the Savilian Professor of Geometry in Oxford at the age of 69. As noted earlier, Sylvester had written about combinatorial systems in 1844 and, in a paper of 1861 [45], he tried to claim priority for the idea of the schoolgirls problem in his own inimitable, but somewhat incomprehensible, way:

> ... in connexion with my researches in combinatorial aggregation ... I had fallen upon the question of forming a heptatic aggregate of triadic synthemes comprising all duads to the base 15, which has since become so well known, and fluttered so many a gentle bosom, under the title of the fifteen school-girls' problem; and it is not improbable that the question, under its existing form, may have originated through channels which can no longer be traced in the oral communications made by myself to my fellow-undergraduates at the University of Cambridge long years before its first appearance, which I believe was in the *Ladies' Diary* for some year which my memory is unable to furnish.

Kirkman quickly dismissed these claims [27]:

> My distinguished friend Professor Sylvester ... volunteers en passant an hypothesis as to the possible origin of this noted puzzle under its existing form. No man can doubt, after reading his words, that he was in possession of the property in question of the number 15 when he was an Undergraduate at Cambridge. But the difficulty of tracing the origin of the puzzle, from my own brains to the fountain named at that university, is considerably enhanced by the fact that, when I proposed the question in 1849, I had never had the pleasure of seeing either Cambridge or Professor Sylvester.

Then, after citing his own paper, Kirkman concluded:

> No other account of it has, so far as I know, been published in print except this guess of Prof. Sylvester's in 1861.

Sylvester also proposed that:

> It were much to be desired that some one would endeavour to collect and collate the various solutions that have been given of the noted 15-school-girl problem by Messrs Kirkman, ..., Moses Ansted [presumably Robert Anstice] ..., by Messrs Cayley and Spottiswoode, ..., and Professor Pierce [Peirce], the latest and probably the best ....

This was eventually done in a paper of F. N. Cole [13] in 1922. It turns out that there are essentially seven different solutions of the schoolgirls problem. A bibliography of early papers relating to the problem was given by Eckenstein [16].

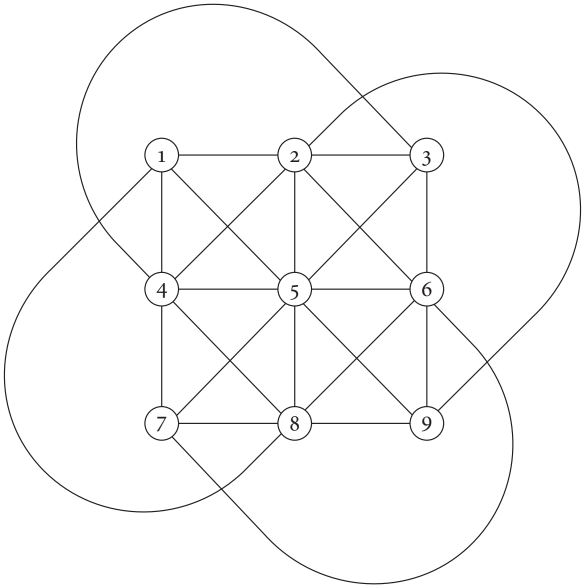
But Sylvester did make one significant contribution to the subject. As reported by Cayley [11] in 1850, Sylvester noticed that there are $C(15, 3) = 455 = 13 \times 35$ possible triples of schoolgirls, and asked whether it is possible to produce thirteen disjoint solutions to the schoolgirls problem, so that each of these 455 triples occurs just once in the quarter-year (thirteen weeks). In 1850 Kirkman [23] claimed (incorrectly) to have a solution, and the problem remained unsolved for over a hundred years until 1971 when R. Denniston [14] used a computer to construct a solution.

Around the same time, the general schoolgirls problem for larger numbers of schoolgirls, in which $6n + 3$ schoolgirls walk in threes over $3n + 1$ days, was solved by Ray-Chaudhuri and Wilson [40]. In fact, an independent solution had been found a few years earlier by Lu Xia Xi, a schoolteacher from Inner Mongolia. However, the extension of Sylvester's problem for $6n + 3$ schoolgirls, where we are required to split the set of all $C(6n + 3, 3)$ triples into $6n + 1$ disjoint solutions of the schoolgirls problem, remains unsolved to this day.

It is sad that Kirkman's name should be remembered primarily for the schoolgirls problem, because his mathematical papers entitle him to be regarded as the founding father of the general theory of what are now called *block designs*, rather than the author of an amusing puzzle. We now consider his contributions in this area and show how they led to other important discoveries in mathematics. A fuller discussion of Kirkman's life and work can be found in Biggs [7].

Projective planes

Plücker's version of the system $S(9)$, described above, was based on a geometrical configuration of points and lines. A pictorial representation is shown below, with lines corresponding to the blocks of $S(9)$.



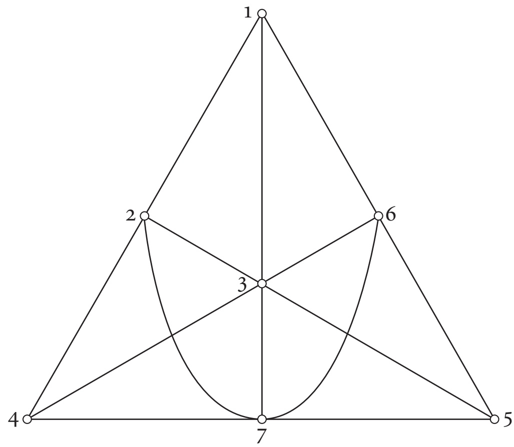
$S(9)$, represented as a geometry.

The geometrical terminology can be applied more generally. The starting point is the observation that in plane geometry *any two points determine a unique line*. This is apparent in our picture of $S(9)$, although some of the lines have had to be 'bent' in order to represent all the relationships correctly. The 'dual' property, that *any two lines meet in a single point*, does not quite hold in ordinary geometry or in $S(9)$, because two lines may be parallel and thus have no point of intersection. However, geometers eventually became dissatisfied with this awkward exception, and invented a new kind of geometry, called *projective geometry*, in which *every* pair of lines meet in exactly one point.

Formally, we define a *projective plane* to be a set of 'points' and a set of 'lines' with the following incidence properties:

- each pair of points is incident with a unique line;
- each pair of lines is incident with a unique point.

(In order to avoid trivial complications, we also assume that not all the points lie on a single line.) It can shown (see below) that a projective plane with a finite number of points must have $s^2 + s + 1$ points and $s^2 + s + 1$ lines, for some positive integer s (called its *order*). Furthermore, each line is incident with exactly $s + 1$ points and each point is incident with exactly $s + 1$ lines. In fact, we have already met such an object with $s = 2$: the system $S(7)$, which can be represented pictorially as follows. Here, too, one of the lines has been bent in order to display the correct incidences.



The seven-point plane (Fano plane).

Kirkman [23] was the first to study finite projective planes. He did not use geometrical language, but he described a method that produces projective planes for each prime value of s. A later paper in a little-known publication, the *Transactions of the Historical Society of Lancashire and Cheshire* [26], contained explicit constructions for projective planes for $s = 2, 3$, and 5, as well as for $s = 4$ and $s = 8$. Significantly, Kirkman failed to obtain a construction in the case $s = 6$, nor was he able to prove that no such configuration can exist.

It was only gradually that geometers became aware that their subject could be studied when the number of points is finite. The idea can be traced back to von Staudt [48], and was developed by the Italian geometer Gino Fano [17] who described finite geometries of various dimensions and, in particular, the above seven-point plane that now bears his name. Shortly afterwards, new algebraic methods were introduced by the American mathematician E. H. Moore, in a paper entitled 'Tactical memoranda I–III' [36]. The crucial fact is that all the standard algebraic operations (those that we now use to define a *field*) can be defined on a finite set of objects if and only if the size of the set is a power of a prime. This observation enabled Moore to set up a method involving latin squares (see Chapter 11) which is equivalent to constructing a projective plane for each prime-power value of s.

In 1906 a paper of Veblen and Bussey [49] continued the work of Fano and Moore. They showed that a famous geometrical theorem of Desargues (on triangles in perspective from a point) holds in any projective plane defined by a finite field. The question of whether there might be projective planes in which Desargues's theorem does not hold was settled almost immediately afterwards, when Veblen and Wedderburn [50] described in some detail a non-Desarguesian plane with $s = 9$.

The Veblen–Wedderburn plane is not defined over a field, but its order s is a prime power. So the question arises:

Is the order of a finite projective plane necessarily a power of a prime?

Recall that Kirkman had failed to construct a plane of order 6, and indeed the impossibility in this case was proved explicitly by MacInnes [33]. (It also follows from results on latin squares described in Chapter 11.) A major step forward was made by Bruck and Ryser [9], who applied some results on integer matrices to the incidence matrix of a projective plane. In this way they proved the non-existence of projective planes of order s for infinitely many values of s, and in particular when $s = 2p$, where p is a prime of the form $4k + 3$.

The outstanding problem that remained was the existence of a plane of order 10, where the complications are immense; indeed, the problem was not resolved until electronic computers could be harnessed to the work. An approach by means of coding theory was suggested by MacWilliams *et al.* [34]. As a result of extensive computer searches, the 'weight enumerator' (which would exist if a plane existed) was completely determined by Lam *et al.* [30] in 1986. Finally, at the end of 1988, Lam and his colleagues announced that a projective plane of order 10 does not exist (see [29] and [31]).

2-designs

In the 1920s the study of combinatorial configurations received another boost, as a result of the work of the statisticians R. A. Fisher and F. Yates. They wanted to design agricultural experiments in such a way that v varieties of wheat (say) are tested in blocks of fixed size k, and each pair of varieties is compared the same number (λ) of times. This led Yates [53] to study systems of 'symmetrized incomplete randomized blocks', which are now known simply as 2-designs.

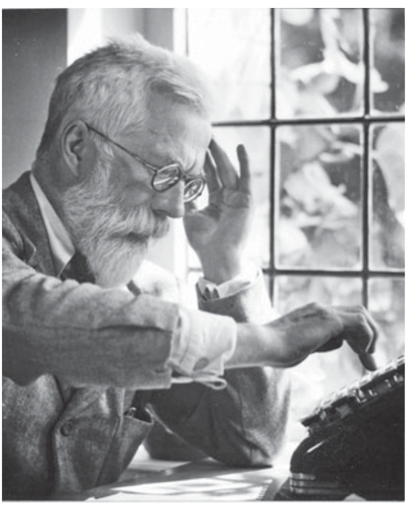
Formally, a 2-*design* with parameters (v, k, λ) consists of a set X of size v, and a collection of subsets of X called *blocks*, each of size k, with the property that each pair of members of X is contained in exactly λ blocks. All of the configurations discussed earlier are 2-designs: the system $S(n)$ is a 2-design with parameters $(n, 3, 1)$, a projective plane of order s is a 2-design with parameters $(s^2 + s + 1, s + 1, 1)$, and there are many other examples. The following table shows the blocks of a 2-design with parameters $(10, 4, 2)$; you may like to check that the pair $\{4, 7\}$ (for example) appears in exactly two blocks.

0 1 2 3	0 1 4 5	0 2 4 6	0 3 7 8	0 5 7 9
0 6 8 9	1 2 7 8	1 3 6 9	1 4 7 9	1 5 6 8
2 3 5 9	2 4 8 9	2 5 6 7	3 4 5 8	3 4 6 7

It is straightforward to prove that, as a consequence of the definition, a 2-design must have another regularity property: *each variety appears the same number of times*, say r. Indeed we have a simple formula for r, and also one for b, the number of blocks:

$$r = (v - 1) \,/\, (k - 1), \;\; b = vr/k;$$

in the above table, you can check that $r = 6$ and $b = 15$. Since the numbers r and b must be integers, these formulas tell us that we cannot just write down any triple of numbers (v, k, λ) and hope to construct a corresponding design; for example, there cannot be a 2-design with parameters $(8, 3, 1)$, because $r = \frac{7}{2}$, which is not an integer.



R. A. Fisher (1890–1962) and Frank Yates (1902–94).

In 1938 Fisher and Yates published a book [19] containing a list of the 2-designs that were known at that time, but many sets of parameters satisfying the elementary conditions were omitted. Around the same time Bose published a long paper [8] describing several new methods of construction. But it was already clear that the above elementary conditions are not sufficient; for example, there is no projective plane of order 6, so the parameters $(43, 7, 1)$ cannot be realized. The problem of finding further necessary conditions was therefore an important one.

In 1940 Fisher [18] showed that the condition $b \geq v$ is a necessary condition for the existence of a 2-design; surprisingly, this is a non-trivial constraint. Chowla and Ryser [12] applied methods similar to those used by Bruck and Ryser for projective planes to designs with $v = b$ and $\lambda \geq 1$, obtaining significant new conditions. In general, it remains an unsolved problem to find a complete set of necessary and sufficient conditions for the existence of a 2-design with parameters (v, k, λ), although R. M. Wilson [51] has shown that the elementary conditions given above are sufficient for existence, provided that v is large enough.

t-designs, for $t \geq 3$

The above designs all satisfy the requirement that every set of two points has a certain property. It is easy to extend the definition to sets of t points, where $t \geq 3$. For example, the following table shows a collection of subsets of size 4 of the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$ with the property that each set of size 3 occurs in exactly one of the blocks.

1235	1248	1267	1346	1378	1457	1568
2347	2368	2456	2578	3458	3567	4678

Generally, we define a *t-design* with parameters (v, k, λ) to be a collection of subsets (blocks) of size k of a set of size v, with the property that every set of size t occurs in exactly λ blocks. (In order to avoid trivial complications, we assume that the blocks do not comprise all the subsets of size k, and that no block appears more than once.) The above figure shows a 3-design with parameters $(8, 4, 1)$, first discovered by Kirkman [25]; in fact, in his paper he gave a general construction for 3-designs with parameters $(2^n, 4, 1)$, for any value of n.

In spite of this early advance, and a few others, it soon became apparent that 3-designs are not plentiful. The situation regarding designs with $t > 3$ seemed even worse. In 1869, a 4-design with parameters $(11, 5, 1)$ was constructed by Lea [32], and Barrau [6] later showed how it could be extended to form a 5-design with parameters $(12, 6, 1)$. These two designs can be obtained from some remarkable finite groups that were first constructed by Claude-Louis Mathieu in the 1860s, but the connection was not properly recognized until much later, when the designs were studied extensively by Carmichael [10] and Witt [52]. These authors also studied two other remarkable designs obtained from Mathieu's groups – a 4-design with parameters $(23, 7, 1)$ and a 5-design with parameters $(24, 8, 1)$. Such designs are interesting not only because of their scarcity, but also because they form the basis for several other important mathematical discoveries; for example, Golay [20] used them to construct two perfect codes. These objects are very unusual – indeed, it has been shown that there can be none besides those that are currently known. For a readable account of this exciting area of mathematics, see Thompson [47].

For many years no new 4-designs or 5-designs were found, and this led to speculation that no others exist, and that for $t > 5$ there are no t-designs whatever. That state of affairs would have been remarkable, implying a deep prejudice against high levels of regularity within a very basic part of finite mathematics. However, in due course some new 4-designs and 5-designs were discovered, by Alltop [2], Assmus and Mattson [5], and Denniston [15]. The suspicion that $t = 5$ might be an absolute limit was then dispelled when Magliveras and Leavitt [35] found several 6-designs with parameters $(33, 8, 36)$. Soon afterwards, Teirlinck [46] surprised the mathematical community by showing that t-designs can be constructed for all values of t. His designs are enormous, and no practical applications have yet been found for them. Nevertheless, there is still considerable interest in the explicit construction of designs with $t \geq 4$.

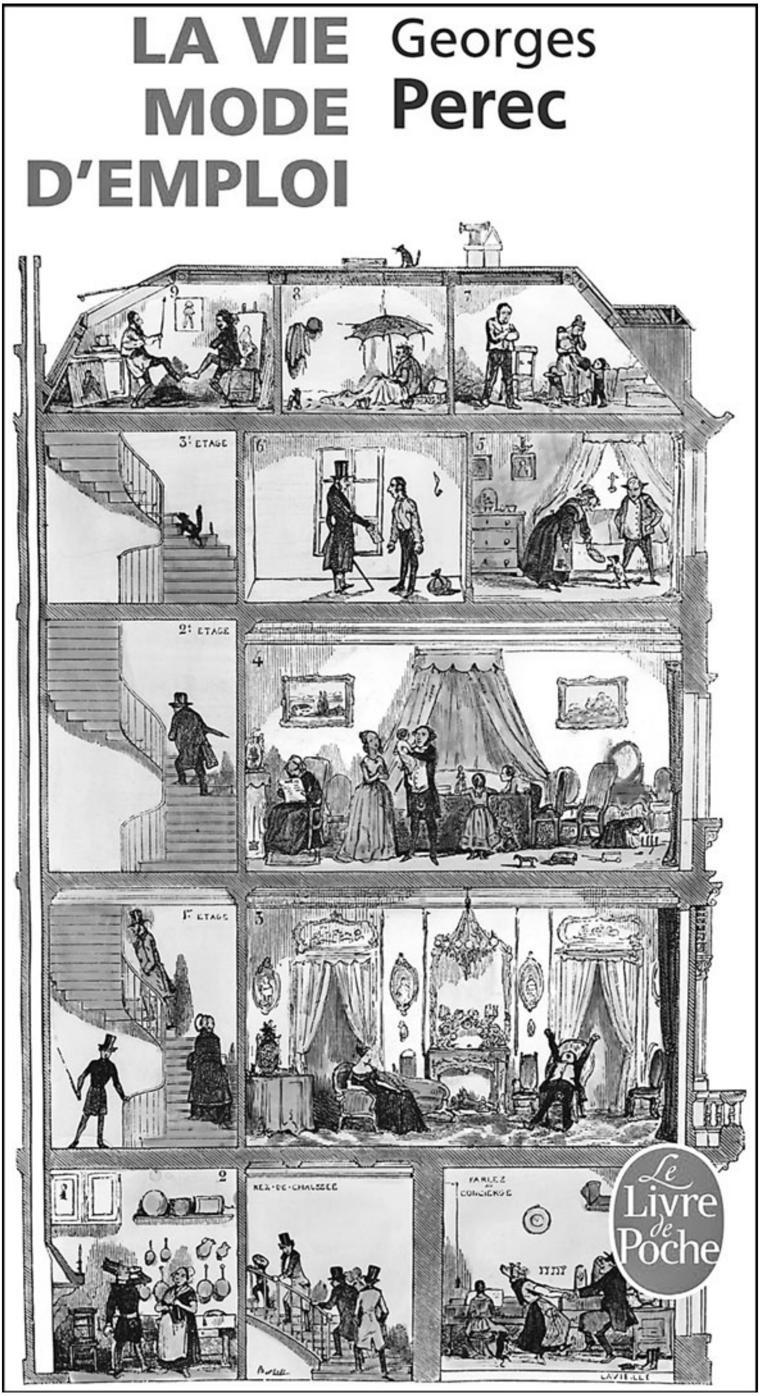
References

1. Earlier versions of parts of this chapter appear in N. L. Biggs, E. K. Lloyd, and R. J. Wilson, The history of combinatorics, *Handbook of Combinatorics* (ed. R. Graham *et al.*), Elsevier (1995), Chapter 44, and also in the second author's paper, The early history of block designs, *Rendiconti del Seminario Matematico di Messina, Serie II, Tomo XXV*, N9 (2003), 267–76, and in his chapter on 'Combinatorics' in R. Flood, A. Rice, and R. Wilson (eds), *Mathematics in Victorian Britain*, Oxford

University Press (2011), 376–95. For useful information on recent results, see C. J. Colbourn and J. H. Dinitz (eds), *The Handbook of Combinatorial Designs*, Chapman and Hall/CRC (second edn, 2006) and its on-line updates.
2. W. O. Alltop, An infinite class of 4-designs, *J. Combin. Theory* 6 (1969), 320–2.
3. Anonymous, [Poem on the problem of the fifteen young ladies], Editorial note by W. J. C. Miller, Solution to Problem 2905, *Math. Questions and their Solutions from the Educational Times* 13 (1870), 79.
4. R. R. Anstice, On a problem in combinations, *Cambridge and Dublin Math. J.* 7 (1852), 279–92.
5. E. F. Assmus and H. F. Mattson, New 5-designs, *J. Combin. Theory* 6 (1969), 122–51.
6. J. A. Barrau, Over de combinatorische opgave van Steiner, *Kon. Akad. Wetensch. Amst. Verslag Wis. en Natuurk. Afd.* 17 (1908), 318–26 = On the combinatory problem of Steiner, *Kon. Akad. Wetensch. Amst. Proc. Sect. Sci.* 11 (1908), 352–60.
7. N. L. Biggs, T. P. Kirkman, mathematician, *Bull. London Math. Soc.* 13 (1981), 97–120.
8. R. C. Bose, On the construction of balanced incomplete block designs, *Ann. Eugenics* 9 (1939), 353–99.
9. R. H. Bruck and H. J. Ryser, The non-existence of certain finite projective planes, *Canad. J. Math.* 1 (1949), 88–93.
10. R. D. Carmichael, Tactical configurations of rank two, *Amer. J. Math.* 53 (1931), 217–40.
11. A. Cayley, On the triadic arrangements of seven and fifteen things, *Philos. Mag. (3)* 37 (1850), 50–3.
12. S. Chowla and H. J. Ryser, Combinatorial problems, *Canad. J. Math.* 2 (1950), 93–9.
13. F. N. Cole, Kirkman parades, *Bull. Amer. Math. Soc.* 28 (1922), 435–7.
14. R. H. F. Denniston, Sylvester's problem of the 15 schoolgirls, *Discrete Math.* 9 (1974), 229–33.
15. R. H. F. Denniston, Some new 5-designs, *Bull. London Math. Soc.* 8 (1976), 263–7.
16. O. Eckenstein, Bibliography of Kirkman's schoolgirl problem, *Messenger of Math.* 41 (1911), 33–36.
17. G. Fano, Sui postulati fondamentali della geometria proiectiva, *Giornale di Matematiche* 30 (1892), 106–32.
18. R. A. Fisher, An examination of the different possible solutions of a problem in incomplete blocks, *Ann. Eugenics* 10 (1940), 52–75.
19. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd (1938).
20. M. J. E. Golay, Notes on digital coding, *Proc. I. R. E.* 37 (1949), 657.
21. T. P. Kirkman, On a problem in combinations, *Cambridge and Dublin Math. J.* 2 (1847), 191–204.
22. T. P. Kirkman, Query VI, *Lady's and Gentleman's Diary* 147 (1850), 48.
23. T. P. Kirkman, Note on an unanswered prize question, *Cambridge and Dublin Math. J.* 5 (1850), 255–62.
24. T. P. Kirkman, Theorems in the doctrine of combinations, *Philos. Mag. (4)* 4 (1852), 209.

25. T. P. Kirkman, Theorems on combinations, *Cambridge and Dublin Math. J.* 8 (1853), 38–45.
26. T. P. Kirkman, On the perfect r-partitions of $r^2 - r + 1$, *Trans. Historical Society of Lancashire and Cheshire* 9 (1857), 127–42.
27. T. P. Kirkman, On the puzzle of the fifteen young ladies, *London, Edinburgh & Dublin Philos. Mag. and J. Sci.* 23 (1862), 198–204.
28. T. P. Kirkman, Appendix I: Solution of Questions 8886 and 9009, *Math. Questions and Solutions from the Educational Times* 47 (1887), 125–7.
29. C. W. H. Lam, The search for a finite projective plane of order 10, *Amer. Math. Monthly* 98 (1991), 305–18.
30. C. W. H. Lam, L. Thiel, and S. Swiercz, The non-existence of code words of weight 16 in a projective plane of order 10, *J. Combin. Theory (A)* 42 (1986), 207–14.
31. C. W. H. Lam, L. Thiel, and S. Swiercz, The non-existence of finite projective planes of order 10, *Canad. J. Math.* 61 (1989), 1117–23.
32. W. Lea, Solution to question 2244, *Math. Questions and Solutions from the Educational Times* 9 (1869), 35–6.
33. C. R. MacInnes, Finite planes with less than eight points on a line, *Amer. Math. Monthly* 14 (1907), 171–4.
34. F. J. MacWilliams, N. J. A. Sloane, and J. W. Thompson, On the existence of a projective plane of order 10, *J. Combin. Theory (A)* 14 (1973), 66–78.
35. S. S. Magliveras and D. W. Leavitt, Simple 6-(33, 8, 36) designs from $P\Gamma L_2(32)$, *Computational Group Theory* (ed. M. D. Atkinson), Academic Press (1984), 337–51.
36. E. H. Moore, Tactical memoranda I–III, *Amer. J. Math.* 18 (1896), 264–303.
37. B. Peirce, Cyclic solutions of the school-girl puzzle, *Astron. J.* VI (No. 6) (1860), 169–74.
38. J. Plücker, *System der analytischen Geometrie, auf neue Betrachtungsweisen gegründet, und insbesondere eine ausführliche Theorie der Curven dritter Ordnung enthalend*, Duncker und Humboldt, Berlin (1835).
39. J. Plücker, *Theorie der algebraischen Curven, gegründet auf eine neue Behandlungensweise der analytischen Geometrie*, Marcus, Bonn (1839).
40. D. K. Ray-Chaudhuri and R. M. Wilson, Solution of Kirkman's schoolgirl problem, *Combinatorics, AMS Proc. Symp. Pure Math.* 9 (1971), 187–203.
41. M. Reiss, Über eine Steinersche combinatorische Aufgabe, *Journal für die reine und angewandte Mathematik* 56 (1859), 326–44.
42. W. Spottiswoode, On a problem in combinatorial analysis, *Philos. Mag. (4)* 4 (1852), 349–54.
43. J. Steiner, Combinatorische Aufgabe, *Journal für die reine und angewandte Mathematik* 45 (1853), 181–2.
44. J. J. Sylvester, Elementary researches in the analysis of combinatorial aggregation, *Philos. Mag. (3)* 24 (1844), 285–96.
45. J. J. Sylvester, Note on the historical origin of the unsymmetrical six-valued function of six letters, *Philos. Mag. (4)* 21 (1861), 369–77.

46. L. Teirlinck, Non-trivial t-designs without repeated blocks exist for all t, *Discrete Math*. 65 (1987), 301–11.
47. T. M. Thompson, *From Error Correcting Codes through Sphere Packings to Simple Groups*, Carus Mathematical Monographs 21, Mathematical Association of America (1983).
48. K. G. C. von Staudt, *Beiträge zur Geometrie der Lage* (3 vols.), Korn'schen Buchhandlung, Nürnberg (1856).
49. O. Veblen and W. H. Bussey, Finite projective geometries, *Trans. Amer. Math. Soc*. 7 (1906), 241–59.
50. O. Veblen and J. H. M. Wedderburn, Non-Desarguesian and non-Pascalian geometries, *Trans. Amer. Math. Soc*. 8 (1907), 379–88.
51. R. M. Wilson, An existence theory for pairwise balanced designs, III: Proof of the existence conjecture, *J. Combin. Theory (A)* 18 (1975), 71–9.
52. E. Witt, Über Steinersche Systeme, *Abh. Math. Sem. Hamburg* 12 (1938), 265–75.
53. F. Yates, Incomplete randomized blocks, *Ann. Eugenics* 7 (1936), 121–40 (reprinted in *Experimental Design: Selected Papers of Frank Yates* (ed. W. G. Cochran *et al*.), Griffin (1970)).



The cover of Georges Perec's book *La Vie Mode d'Emploi*. This book has one hundred chapters that correspond to a knight's tour through a 10 × 10 array consisting of two superimposed orthogonal latin squares of order 10.

CHAPTER 11

Latin squares

LARS DØVLING ANDERSEN

A latin square of order n is an $n \times n$ array with entries from a set of n symbols, arranged in such a way that each symbol appears exactly once in each row and exactly once in each column. From this simple starting point, the theory of latin squares has developed into an interesting discipline in its own right, as well as an important tool in design theory in general.

Introduction

1	2	3
2	3	1
3	1	2

IGNIS			
IGNIS	AER	AQVA	TERRA
AER	IGNIS	TERRA	AQVA
AQVA	TERRA	IGNIS	AER
TERRA	AQVA	AER	IGNIS

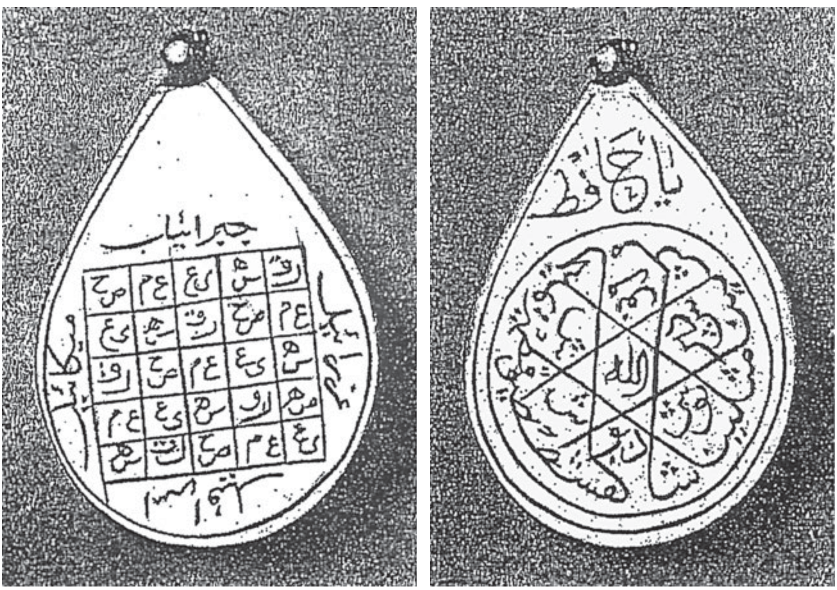
فلان	الرحيم	الرحمن	الله	بسم
بسم	فلان	الرحيم	الرحمن	الله
الله	بسم	فلان	الرحيم	الرحمن
الرحمن	الله	بسم	فلان	الرحيم
الرحيم	الرحمن	الله	بسم	فلان

Latin squares of orders 3, 4, and 5.

Consider the three latin squares above. The 3×3 latin square is a *cyclic* latin square because the symbols appear in the same cyclic order in each row and column. It is clear that cyclic latin squares exist for all orders n: write the n

symbols in any order in the top row; for the second row shift each symbol one column to the left, placing the first symbol in the last column; then continue in this fashion row by row, always shifting by one column to the left. An equivalent way to think of this construction, assuming that the symbols are the integers 1 to n, is that each new row is formed by adding 1 to every number in the previous row modulo n. The two other latin squares above are original examples of early latin squares.

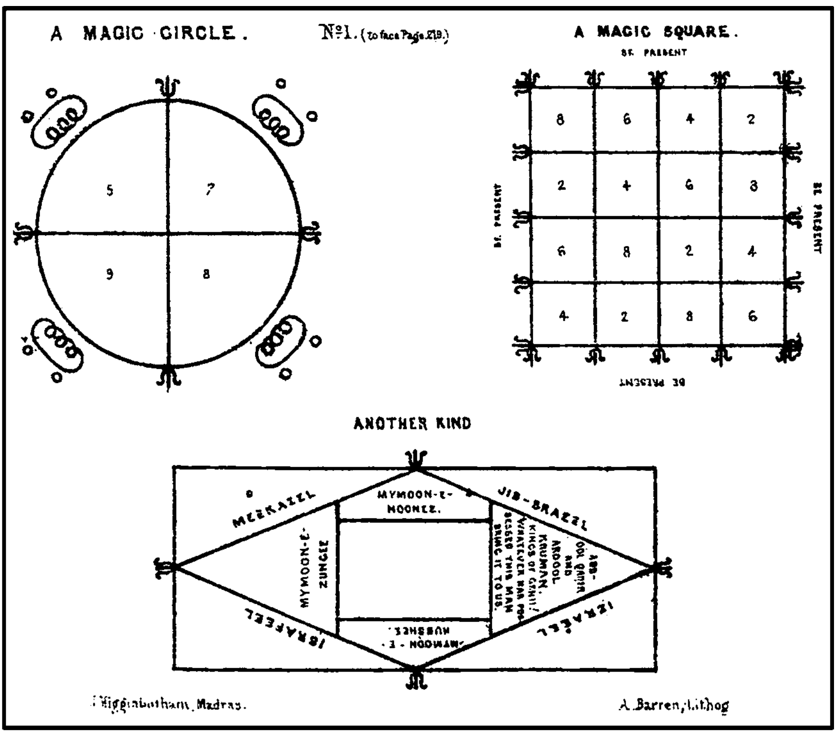
The first known occurrences of latin squares seem to have been in their use on amulets and in rites in certain Arab and Indian communities from perhaps the year 1000: the nature of the sources makes the dating difficult. Most similar amulets contain not a latin square, but a *magic square* – an $n \times n$ array filled with the symbols $1, 2, \ldots, n^2$ for which the sum of the numbers in any row, column, or main diagonal is the same. The latin square amulets, like the magic square ones, were worn to fight evil spirits, show reverence for gods, celebrate the Sun and the planets, etc.; in medieval books on magic and latin squares, they are often framed by fanciful ornamental structures (see [1]).



A silver amulet from Damascus. On one side is a latin square, and on the other side are the names of the Seven Sleepers, who, according to legend, slept in a cave for two hundred years from about the year 250.

But the use of such squares was not restricted to amulets or talismans. The following picture shows how a latin square (here called a magic square) was considered, together with a magic circle, to have powers for casting out devils (see [2]); instructions for their use were given as:

Magic circles, squares, and figures, are sketched in the ground, or on a plank, with various coloured powders, bhubhoot [cow-dung ashes], charcoal, or sundul; and the demoniac being seated in the centre of it, the afsoon [incantation] is read. Around these diagrams are placed various kinds of fruits, flowers, pan-sooparee, sheernee, sometimes sayndhee, taree, nariellee [intoxicating liquors], daroo [ardent spirits], etc. Some sacrifice a lamb in front of the circle, sprinkle blood around it, set up the head in front, placing a lamb upon it, lighted up with a puleeta [charm-wick]; or they merely slay a fowl, and sprinkle its blood around. Some give a rupee or two, according to their means, into the hands of the person possessed by the devil, to deposit therein.



The entries of the latin square, the numbers 2, 4, 6, and 8, were considered throughout the Islamic Orient to have magic powers when occurring together, in which case they could be associated with an often-used magic word. It is significant that they also occur on both diagonals and in the four corners of the square.

Tracing such occurrences of latin squares back in history leads to a famous book, *Shams al-Ma'arif al-Kubra* (The Sun of Great Knowledge) [3], which was written by Aḥmad ibn 'Alī ibn Yūsuf al-Būnī, an Arab Sufi believed to have died in 1225. This book contains many latin squares (in addition to many more magic squares), including the above 5×5 square (the first latin square of the book, and thus possibly the oldest known latin square) and a description of a talisman containing seven latin squares, each associated with both a weekday and a planet.

حرف الظاء للمشتري وله يوم الخميس

ظ	ث	ج	ف	خ	ش	ظ
ج	ف	خ	ش	ظ	ز	ث
خ	ش	ظ	ز	ث	ج	ف
ظ	ز	ث	ج	ف	خ	ش
ث	ج	ف	خ	ش	ظ	ز
ف	خ	ش	ظ	ز	ث	ج
ش	ظ	ز	ث	ج	ف	خ

One of seven latin squares of a talisman from al-Būnī's book, associated with Thursday and Jupiter. The first entry is in error, and although the structure of each square is obvious, only three of them are correct.

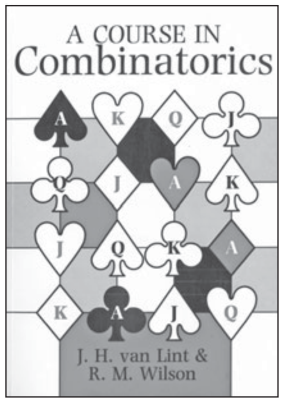
The latin squares of al-Būnī seem to serve two purposes: first, they have certain magic powers in their own right, some of them being related to a specific planet; and secondly, and mathematically very interestingly, they seem to be crucial in the construction of magic squares. In fact, there is strong evidence [4] that al-Būnī and other early Arab authors knew of methods for this that were later taken up by 18th-century mathematicians, among them Leonhard Euler.

Likewise, a book from 1356 on Indian mathematics [5] contains latin squares with a clear focus on their use in certain constructions of magic squares. Thus, Hindu mathematics also anticipated what Euler later formalized and developed, as we see below. In the succeeding centuries, more texts contained such applications.

In the 13th century the Catalan mystic and philosopher Ramon Llull constructed latin squares in his efforts to explain the world by combinatorial means; the 4×4 square at the beginning of this chapter is taken from a larger drawing of Llull's (see Chapter 5).

An ancient card problem asks for the sixteen court cards in an ordinary pack of playing cards to be arranged in a 4×4 array so that each row, each column, and each main diagonal contains an ace, a king, a queen, and a jack, all of different suits; thus, both the suits and the values form latin squares. From the early 18th century or earlier, this problem featured regularly in collections of mathematical problems of a recreational nature [6].

AS DE CŒUR	ROI DE TRÈFLE	DAME DE CARREAU	VALET DE PIQUE
VALET DE CARREAU	DAME DE PIQUE	ROI DE CŒUR	AS DE TRÈFLE
ROI DE PIQUE	AS DE CARREAU	VALET DE TRÈFLE	DAME DE CŒUR
DAME DE TRÈFLE	VALET DE CŒUR	AS DE PIQUE	ROI DE CARREAU



Two solutions to the card puzzle.

A latin square also appears on a brass plate in St Mawgan Church in Cornwall. It takes the form of a poem commemorating a certain Hanniball Basset, who died in 1709:

Shall	wee	all	dye
Wee	shall	dye	all
All	dye	shall	we
Dye	all	wee	shall

Like much of design theory, latin squares have applications in statistics, in experimental design. The earliest known example is by the French agricultural researcher François Cretté de Palluel, who presented a paper to the Royal Agricultural Society of Paris on 31 July 1788 [7]. His purpose was to show that one might just as well feed sheep on root vegetables during winter – this was much cheaper and easier than the normal diet of corn and hay – and he described an experiment of feeding sixteen sheep with different diets and comparing their weight gains.

Although no latin square appears in his published paper, the layout of his sheep experiment amounted to a 4 × 4 latin square with four breeds of sheep as rows, four different diets as columns, and four different slaughtering times as symbols.

EXPERIMENT UPON FATTING SHEEP, AND THEIR INCREASE FROM MONTH TO MONTH.

Sixteen ſheep, of the ſame age, of four different breeds, were picked out of my flock, viz. four the breed of the country, four of Beauce, four of Champagne, and four of Picardy; I weighed them alive, and marked each with a number; I divided them into four lots, and fed them on four different ſorts of food, as under:

Food.	No.	Breeds.	Weights at different Periods.—1788. Jan. 20.	Feb. 20.	Mar. 20.	April 20.	May 20.	Increaſe each Month. 1ſt M.	2d M.	3d M.	4th M.	To'al incr. which each food has produced upon four ſheep.
Potatoes,	1	Iſle de France,	69¾ lb.	79¼ lb.	—	—	—	10 lb.	lb.	lb.	lb.	70 lb.
	2	Beauce,	70¼	82½	90¼ lb.	93 lb.	95 lb.	11½	7½	2¾	2	
	3	Champagne,	69¼	83	82½	84	—	13¾	loſs ½	1½	—	
	4	Picardy,	88	95	101	—	—	15	6	—	—	
								50½	13¾	4¾	2	
Turnips,	5	Iſle de France,	69	86	87	—	—	17	1	—	—	67½
	6	Beauce,	71	86	—	—	—	15	—	—	—	
	7	Champagne,	68½	78¼	82½	84	84½	10	4	1¼	¼	
	8	Picardy,	79	95¼	97½	97½	—	16½	2	—	—	
								58½	7	1½	½	
Beets,	9	Iſle de France,	72	83¼	90¾	94	—	11¼	7¼	3¼	—	71
	10	Beauce,	70¼	80¼	86	—	—	10	5¼	—	—	
	11	Champagne,	77¼	90½	—	—	—	13¼	—	—	—	
	12	Picardy,	80	93½	98½	100½	101	13½	5	1½	½	
								48	17½	5	½	
Oats, Barley, and grey peas.	13	Iſle de France,	74	91	95½	102	106	17	4½	6½	4	92½
	14	Beauce,	73½	84¼	91½	96	—	10¼	7¼	4½	—	
	15	Champagne,	71	86¼	93	—	—	15¼	6¼	—	—	
	16	Picardy,	71	87	—	—	—	16	—	—	—	
								59	18½	11	4	

Observation. The increaſe of theſe ſheep, during the firſt month, being ſo much more conſiderable than in the following months, muſt be attributed to this cauſe, that lean cattle put up to fatten, eat greedily until they are cloyed, which only fills them, without much increaſing their fleſh; but, on the contrary, the increaſe produced in the enſuing months, although apparently leſs, turns all to profit in fleſh and tallow.

A page from the English translation of Cretté de Palluel's 1788 paper.

More recently, latin squares have featured in a number of artistic situations. In Gonville and Caius College, Cambridge, there is a stained-glass window commemorating their former student R. A. Fisher. It is a colourful 7×7 latin square by the artist Maria McClafferty; another window, by the same artist, commemorates John Venn. It also happens that some postage stamps are issued in square sheets of n^2 stamps, arranged as a latin square (see [8]).

The latin square window at Gonville and Caius College.

Euler and latin squares

Shortly before de Palluel's paper appeared, latin squares were introduced to the mathematical community by Leonhard Euler: he gave them their name, and seems to have been the first to define them using mathematical terminology and to investigate their properties mathematically. Although he had known and used them a little earlier, he first published latin squares in a paper that began with his famous 'thirty-six officers problem', presented to the St Petersburg Academy of Sciences in 1779 and published in 1782. He thereby launched a more complicated concept: *orthogonal Latin squares*.

> 1. Une question fort curieuse, qui a exercé pendant quelque temps la sagacité de bien du monde, m'a engagé à faire les recherches suivantes, qui semblent ouvrir une nouvelle carrière dans l'Analyse et en particulier dans la doctrine des combinaisons. Cette question rouloit sur une assemblée de 36 officiers, de six différens grades et tirés de six régimens différens, qu'il s'agissoit de ranger dans un quarré de manière que sur chaque ligne, tant horizontale que verticale, il se trouvât six officiers tant de différens caractères que de régimens différens. Or, après toutes les peines qu'on s'est données pour résoudre ce problème, on a été obligé de reconnoître qu'un tel arrangement est absolument impossible, quoiqu'on ne puisse pas en donner de démonstration rigoureuse.

The introductory question in Euler's paper.

Euler explained that the contents of the paper were inspired by a strange question about a collection of thirty-six officers, of six different ranks and from six different regiments, who should line up in a square in such a way that in each line, both horizontal and vertical, were six officers both of different ranks and from different regiments. He added that he had come to realize that such an arrangement was impossible, but could not prove it.

Euler clarified the officers problem by denoting the regiments by Latin letters, a, b, c, d, e, f, and the ranks by Greek letters, $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$. He explained that the task was to arrange the thirty-six pairings of a Latin and a Greek letter in a 6×6 array so that each row and column contained each Latin and Greek letter just once: the 36 cells would then contain every possible pairing of a Latin symbol and a Greek one.

He stated that since he had not been able to solve this problem, he would generalize it to pairs from n Latin and n Greek letters for an arbitrary whole number n. He then worked with the numbers $1, 2, \ldots, n$, instead of Latin and Greek letters, and introduced the concept of a latin square: it was called a latin square because its numbers could be Latin letters in what might have a counterpart with Greek letters satisfying the all-pairings property!

1^1	2^6	3^4	4^3	5^7	6^5	7^2
2^2	3^7	1^5	5^4	4^1	7^6	6^3
3^3	6^1	5^6	7^5	1^2	4^7	2^4
4^4	5^2	6^7	1^6	7^3	2^1	3^5
5^5	1^3	7^1	2^7	6^4	3^2	4^6
6^6	7^4	4^2	3^1	2^5	5^3	1^7
7^7	4^5	2^3	6^2	3^6	1^4	5^1

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

The first latin square to go by this name. On the left is a square of pairs: each number from 1 to 7 occurs exactly once in each row and each column, both in line and as an exponent, and each ordered pair occurs exactly once. Referring to his earlier usage of Latin and Greek letters, Euler called the in-line numbers 'Latin numbers' and the exponents 'Greek numbers'. He then dropped the exponents, obtaining the *latin square* on the right.

We now say that two latin squares of the same order are *orthogonal* if they have the property that whenever two places have the same entry in one square, then they have distinct entries in the other; it follows that if the two squares are superimposed, then the n^2 cells contain each possible pairing of a symbol from the first square and one from the second. Thus, the officers problem asks for two orthogonal latin squares of order 6. In the figure above, the square to the left shows two orthogonal latin squares of order 7, showing that Euler could solve the corresponding forty-nine officers problem. In the same way, if we ignore the condition on the diagonals, the above card problem asks for two orthogonal latin squares of order 4, and the solution presented shows that such orthogonal latin squares exist.

Euler's 1782 paper was called 'Recherches sur une nouvelle espèce de quarrés magiques' (Research on a new kind of magic square). It contains a few references to magic squares, showing how orthogonal latin squares can be used for constructing them. However, there is more evidence that Euler came to consider latin squares through an interest in magic squares. These were well known at the time, and he seems to have worked on them at an early age and returned to them fifty years later. In his mathematical notebooks there is a brief piece on magic squares, believed to be from 1726, and in 1776 he presented a long paper on the topic to the Academy of Sciences of St Petersburg. Both were published posthumously, and both were entitled 'De quadratis magicis' (On magic squares).

We find the evidence in his 1776 paper. In this paper Euler's first construction is of a 3×3 magic square. He explained that he needed the Latin letters a, b, c and the Greek letters α, β, γ to be given numerical values. He then displayed the square

$$\begin{matrix} a & b & c \\ b & c & a \\ c & a & b \end{matrix}$$

with Latin letters and noted that each row and column has the same sum. Thus, this square has constant row, column, and diagonal sums if and only if $a + b + c = 3c$ – that is, $2c = a + b$.

He wished to obtain his magic square by adding to the above square a similar one with Greek letters:

$$\begin{matrix} \gamma & \beta & \alpha \\ \alpha & \gamma & \beta \\ \beta & \alpha & \gamma, \end{matrix}$$

giving the square array

$$\begin{matrix} a+\gamma & b+\beta & c+\alpha \\ b+\alpha & c+\gamma & a+\beta \\ c+\beta & a+\alpha & b+\gamma. \end{matrix}$$

This method requires that the same two letters are not added twice, so the square obtained by combining the first two squares,

$$\begin{matrix} a\gamma & b\beta & c\alpha \\ b\alpha & c\gamma & a\beta \\ c\beta & a\alpha & b\gamma, \end{matrix}$$

also displayed by Euler, must have the property that all nine pairings of letters occurring are distinct. Letting a, c, b and α, γ, β be arithmetic progressions, one with difference 1 and the other with difference 3 (the size of the square), we readily produce a magic square. Euler put $(a, c, b) = (0, 3, 6)$ and $(\alpha, \gamma, \beta) = (1, 2, 3)$ and obtained the magic square

$$\begin{matrix} 0 & 6 & 3 \\ 6 & 3 & 0 \\ 3 & 0 & 6 \end{matrix} \quad + \quad \begin{matrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{matrix} \quad = \quad \begin{matrix} 2 & 9 & 4 \\ 7 & 5 & 3 \\ 6 & 1 & 8. \end{matrix}$$

What he has done here, without using this terminology, is to find two *orthogonal* 3×3 latin squares and use them to construct a 3×3 magic square.

Neither in his 1776 paper, nor in his 1782 paper, did Euler use the phrase 'orthogonal latin squares'. He also never used the term *Graeco-Latin square* for a pair of orthogonal latin squares, a term frequently attributed to him – and certainly not the name *Euler square*, also used later.

Euler then discussed the same method for 4×4 squares. Here he found two orthogonal Latin squares, both with the property that each diagonal also contains all the symbols. The combined square that he displayed is

$$\begin{array}{cccc} a\alpha & b\delta & c\beta & d\gamma \\ d\beta & c\gamma & b\alpha & a\delta \\ b\gamma & a\beta & d\delta & c\alpha \\ c\delta & d\alpha & a\gamma & b\beta. \end{array}$$

This has the sum property for a magic square, no matter what values are given to the letters. Euler noted that if a, b, c, d are assigned the values $0, 4, 8, 12$ (in any order), and if $\alpha, \beta, \gamma, \delta$ are assigned the values $1, 2, 3, 4$ (in any order), then all the integers from 1 to 16 are obtained as pair sums and a magic square is obtained. He was thereby able to obtain $4! \times 4! = 576$ distinct magic squares of side 4; in fact, other assignments are possible, giving further magic squares. Euler also found orthogonal latin squares of order 5, each with distinct symbols on both diagonals, thereby obtaining $5! \times 5! = 14\,400$ distinct magic squares of order 5.

We now return to the old ways of making magic squares by means of latin squares, mentioned earlier. A Hindu method for combining squares can be described as follows [9]:

- take a cyclic latin square of odd order with the numbers in the middle column appearing in natural order;
- place next to it a copy where all the entries are multiplied by the order of the square;
- flip the second square on top of the first, as if closing a book, and add the numbers in corresponding cells.

The result is a magic square formed from numbers beginning with one more than the order of the square; subtracting the order from each entry then gives an ordinary magic square. For example, the Hindu method can be used to create a magic square of order 5, as follows:

$$
\begin{array}{ccccc|ccccc}
4 & 5 & 1 & 2 & 3 & 20 & 25 & 5 & 10 & 15 \\
5 & 1 & 2 & 3 & 4 & 25 & 5 & 10 & 15 & 20 \\
1 & 2 & 3 & 4 & 5 & 5 & 10 & 15 & 20 & 25 \\
2 & 3 & 4 & 5 & 1 & 10 & 15 & 20 & 25 & 5 \\
3 & 4 & 5 & 1 & 2 & 15 & 20 & 25 & 5 & 10
\end{array}
$$

$$
\longrightarrow
\begin{array}{ccccc}
19 & 15 & 6 & 27 & 23 \\
25 & 16 & 12 & 8 & 29 \\
26 & 22 & 18 & 14 & 10 \\
7 & 28 & 24 & 20 & 11 \\
13 & 9 & 30 & 21 & 17
\end{array}
\longrightarrow
\begin{array}{ccccc}
14 & 10 & 1 & 22 & 18 \\
20 & 11 & 7 & 3 & 24 \\
21 & 17 & 13 & 9 & 5 \\
2 & 23 & 19 & 15 & 6 \\
8 & 4 & 25 & 16 & 12
\end{array}
$$

The construction works because the latin square obtained from a cyclic one of odd order by interchanging left and right is orthogonal to the original square. For two orthogonal latin squares obtained in this way, the method is exactly the same as that of Euler.

An analysis of the magic squares in al-Būnī's book shows that these could have been formed by a similar method and thus be based on orthogonal latin squares [10]. The method was taken up in later writings about magic squares before Euler, and books by de la Loubère (1691) and Poignard (1704) and a paper by la Hire (1705) led in 1710 to Joseph Sauveur publishing a text containing many orthogonal latin squares [11], as he had conceived the same idea as Euler for constructing magic squares.

Sauveur used upper and lower case Latin letters where Euler later used Latin and Greek letters; in the next section we present an illustration from his paper where Greek letters appear as well, because here he needed three alphabets (de la Loubère had already introduced Greek letters in his work on magic squares). But although Sauveur discussed orthogonal latin squares in an abstract way over sixty years before Euler did, he did not extract the idea of a single latin square: nowhere in his paper is there an ordinary latin square with a single symbol in each cell. On the other hand, the book by Poignard contained many such squares, of orders 3 –10, 12, and 16, and with numbers as entries [12].

We should also note that a pair of orthogonal latin squares of order 9 appeared in the early 18th century in Korea, in a publication in Chinese by Choe Sŏk-chŏng [13]. These can be seen as being based on one latin square of order 9 consisting of nine 3×3 subsquares, and the other 9×9 square being orthogonal to the first and obtained from it by interchanging left and right as in the Hindu case described above.

In the final section of his 1776 paper Euler considered magic squares of order 6, but here, as we have seen, he was not able to produce a pair of orthogonal latin squares. Instead, he presented a similar construction (which he also introduced for 4×4 squares) where the entries are still sums of a Latin letter and a Greek letter, but where the squares formed by each type individually are no longer latin squares (in fact, there was an error in his construction). Such methods for constructing magic squares, by adding the entries of two auxiliary squares that are not both latin squares, were also known to the earlier authors mentioned above.

On 8 March 1779 Euler publicly announced such considerations in his St Petersburg lecture on the officers problem, and in his 1782 paper he proceeded to consider orthogonal latin squares in general. In discussing these, Euler introduced another important concept for latin squares: a *transversal* (in Euler's paper, 'une formule directrice') in a latin square of side n is a set of n distinct entries occurring in distinct rows and distinct columns. He explained that for a latin square of side n to have a latin square orthogonal to it (in modern terminology, an *orthogonal mate*), it must have n mutually disjoint transversals, each corresponding to the occurrences of a particular symbol in the other square. Euler actually stated that the search for transversals was the main object of the paper, but added that he had no method for finding them.

First, he looked for transversals in cyclic latin squares, and he proved that a cyclic latin square of side n has no transversals when n is even; more precisely, he proved only that there is no transversal containing the first entry of the cyclic square, but this generalizes easily. For the proof, he assumed that there is such a transversal, and that its entries are 1 in the first column, a in the second, b in the third, and so on. If these occur in rows $1, \alpha, \beta, \ldots$, then, by the definition of the cyclic square and calculating modulo n, we have

$$a = \alpha + 1, \; b = \beta + 2, \; c = \gamma + 3, \text{ etc.}$$

(Note the ease with which Euler let the distinction between Latin and Greek letters serve a different purpose from that at the outset of the paper.) Since $\{a, b, c, \ldots\} = \{\alpha, \beta, \gamma, \ldots\} = \{2, 3, \ldots, n\}$, he obtained

$$S = a + b + \cdots = (\alpha + 1) + (\beta + 2) + \cdots = S + \tfrac{1}{2}n(n-1),$$

which is true if and only if n is odd. A consequence is that

Cyclic latin squares of even order do not have orthogonal mates.

For odd n, the main diagonal is a transversal, and from the cyclic property it is easy to find a set of n disjoint transversals. Thus,

Cyclic latin squares of odd order have orthogonal mates,

and consequently,

Orthogonal latin squares exist for all odd orders.

Euler noted this, and also gave rules for finding many other transversals from a given one.

In the next three sections of his long paper, Euler investigated the existence of transversals in latin squares built cyclically from cyclic 2×2, 3×3, and 4×4 squares. Along the way, he found new examples of latin squares without transversals. He also proved that

There are orthogonal latin squares of all orders n divisible by 4.

The orthogonal latin squares of odd order based on the diagonal transversals were exactly those relevant to the Hindu method we saw earlier, and the Indians had also been aware of the existence of orthogonal latin squares of orders divisible by 4.

In conclusion, Euler offered the now famous Euler conjecture on orthogonal latin squares:

Conjecture: There is no pair of orthogonal latin squares of side n, for $n \equiv 2$ $(mod\ 4)$

– that is, for $n = 2, 6, 10, 14, 18, \ldots$. Euler knew this to be true for $n = 2$, but (as we shall see in the next section) it took more than a century before Gaston Tarry proved it true for $n = 6$, thereby settling the officers problem in the negative. Almost two centuries elapsed before R. C. Bose, E. T. Parker, and S. Shrikhande proved it to be false for *all* other values of $n \equiv 2 \pmod{4}$.

The final pages of Euler's remarkable paper contained some pointers to future developments. He observed that a pair of orthogonal latin squares of side n can be described by a list of n^2 quadruples, each consisting of a row number, a column number, the number in this position in the first square, and the number in the same position in the second square; this anticipated the later notion of an orthogonal array. Euler realized that the meaning of the positions in the quadruples is interchangeable, so that given one pair of orthogonal latin squares of a given side, there could be twenty-four such pairs – although he knew that these may not all be distinct. He also stated that he considered the problem of enumerating latin squares to be very important, but also very difficult.

As can be seen from this discussion, magic squares do not play a prominent role in Euler's 1782 paper, and yet he chose to put their name in the title. Whatever the reason for this – to attract readers, perhaps, because magic squares were popular – it seems clear that the then unpublished paper from his 1776 lecture contained important preliminary work. It was, however, the concept of 'orthogonality' that was to prove fruitful for future combinatorial advances.

Mutually orthogonal latin squares

As we have seen, Euler's 'officers problem' and his more general conjecture remained unresolved for a long time. The problem seems to have been well known, at least in the late 19th century, when several papers on it appeared. Euler's paper was reissued in 1849, and this may have helped to arouse interest.

In 1900 Tarry [14] proved the conjecture true when $n = 6$, the officers case; this is the earliest preserved proof known, although possibly not the first to be conceived. (Among those publishing false proofs was the Danish mathematician Julius Petersen, whose paper appeared after Tarry's.) Tarry partitioned the set of 6×6 latin squares into seventeen classes and did a case-by-case analysis of these.

Remarkably, there is an 1842 letter [15] from Heinrich Schumacher to Carl Friedrich Gauss informing him that the astronomer Thomas Clausen had solved the question by reducing it to seventeen cases, but there is no known written work from Clausen on this. Schumacher and Gauss discussed Euler's general conjecture without mentioning Euler, writing 'Clausen vermuthet, dass es für jede Zahl von der Form $4n + 2$ unmöglich sei . . . ' (Clausen suspects that for numbers of the form $4n + 2$ it is impossible . . .).

Since Tarry's case-by-case solution appeared, people have continued to look for simpler proofs of the officers case, and a number have been found. Notable examples of these are by Fisher and Yates (1934) [16], Betten (1983) [17], and Stinson (1984) [18].

It is possible for more than two latin squares of the same order to be mutually orthogonal; for example, any two of the following three latin squares are orthogonal:

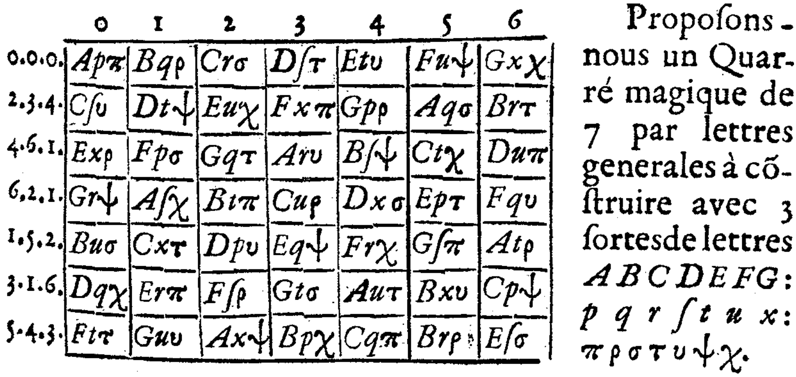
1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

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

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

Such sets of 'mutually orthogonal latin squares' are now referred to as MOLS.

Before further progress on the Euler conjecture was made, H. F. MacNeish (1922) mentioned a possible generalization of Euler's conjecture to larger sets of MOLS (see [19]). For any positive integer n, let $N(n)$ be the maximum number of MOLS of order n; for example, $N(4) = 3$. It is not difficult to see that $N(n) \leq n - 1$ for all $n > 1$, since all the first rows can be assumed to be $1, 2, \ldots, n$, and the latin squares must then have different symbols (other than 1) in the first cell of the second row. In his 1710 paper [11], Sauveur published three MOLS of order 7 without using this terminology; this shows that $N(7) \geq 3$.



	0	1	2	3	4	5	6
0.0.0.	Apπ	Bqρ	Crσ	Dſτ	Etυ	Fuψ	Gxχ
2.3.4.	Cſυ	Dtψ	Euχ	Fxπ	Gpρ	Aqσ	Brτ
4.6.1.	Exρ	Fpσ	Gqτ	Arυ	Bſψ	Ctχ	Duπ
6.2.1.	Grψ	Aſχ	Btπ	Cuρ	Dxσ	Epτ	Fqυ
1.5.2.	Buσ	Cxτ	Dpυ	Eqψ	Frχ	Gſπ	Atρ
3.1.6.	Dqχ	Erπ	Fſρ	Gtσ	Auτ	Bxυ	Cpψ
5.4.3.	Ftτ	Guυ	Axψ	Bpχ	Cqπ	Brρ	Eſσ

Propoſons-nous un Quarré magique de 7 par lettres generales à cõſtruire avec 3 ſortesde lettres *ABCDEFG*: *p q r ſ t u x*: π ρ σ τ υ ψ χ.

Sauveur's three mutually orthogonal latin squares of order 7.

In 1922 MacNeish defined an *Euler square of order n and degree k* to be a square array of n^2 k-tuples with properties corresponding to the entries defining k MOLS; he also said that such an Euler square has *index* (n, k). Thus, Sauveur's diagram shows an Euler square of index $(7, 3)$ consisting of forty-nine triples. He required that $k \leq n - 1$, so was obviously aware of this necessary condition.

MacNeish stated two basic results. The first is that

If q is a prime power, then there exist $q - 1$ *MOLS of order q*

– that is, there exists an Euler square of index $(q, q-1)$.
The second, in present-day language, is that

There is a direct product construction for combining k MOLS of order m and k MOLS of order n to give k MOLS of order mn.

He mentioned that this construction extends a method that is similar to 'the method for combining two magic squares' used by Tarry for index 2.

MacNeish's arguments were not always watertight (his paper also included an erroneous proof of the Euler conjecture) and, as Parker later observed, the construction was 'put on an algebraic foundation by H. B. Mann' (although Moore had done it earlier, and Bose had done it before Mann). Still, MacNeish is usually credited with the following more general result, which follows from the above statements:

If the prime factorization of n is $p_1^{r_1} p_2^{r_2} \cdots p_t^{r_t}$*, then*

$$N(n) \geq \min(p_1^{r_1}, p_2^{r_2}, \ldots, p_t^{r_t}) - 1.$$

MacNeish mentioned the possibility that the bound in this statement is actually the true value of $N(n)$, as is the case when n is a prime power. If $n \equiv 2 \pmod 4$, then one of the prime powers in the product is 2, so the conjectured value for $N(n)$ is 1, in agreement with Euler's conjecture. He did not explicitly state this as a conjecture in his paper, but wrote that the proof of this result 'is a generalization of the Euler problem of the 36 officers which has not been proved'.

For a third of a century, both Euler's conjecture and MacNeish's generalization of it remained open, and there were even published results lending support to both conjectures. But in 1958–59, both conjectures fell. First, Parker discovered four MOLS of order 21, thereby disproving MacNeish's conjecture (his construction yielded other counter-examples as well). Then Bose and Shrikhande constructed two MOLS of order 22 and five MOLS of order 50, and after that Parker found two MOLS of order 10; it was mentioned in these papers that there are infinitely many numbers n for which Euler's conjecture does not hold. Finally, Bose, Shrikhande, and Parker [20] disproved the conjecture for all $n \equiv 2 \pmod 4$, other than $n = 2$ and 6. The methods used by these so-called 'Euler's spoilers' (a phrase coined by Martin Gardner [21]) had a general feature that is so often useful: that of creating new designs from old ones in clever ways.

By this time, the problem had become so famous that on 26 April 1959 its solution was announced on the front page of *The New York Times* in a news story 'Major Mathematical Conjecture Propounded 177 Years Ago Is Disproved'. When two orthogonal latin squares of order 10 were found, the November 1959 cover of the magazine *Scientific American* reproduced a painting by its staff artist Emi Kasai illustrating the squares with colours. In the following year, Mrs Karl Wihtol made a needlepoint rug from the painting.

Two orthogonal latin squares of order 10 as a needlepoint rug.

Moreover, when the French author Georges Perec wrote his masterpiece *La Vie Mode d'Emploi* (Life: A User's Manual) in 1978, he thrust upon himself a number of restrictions of a mathematical nature. One of these was that its one hundred chapters should correspond to a knight's tour through a 10×10 array consisting of two superimposed orthogonal latin squares of order 10. He achieved this by letting the physical locations of the chapters be different parts of a nine-storey building with a basement; the illustration that opens this chapter gives a hint of this.

In 1960 four consecutive papers in the *Canadian Journal of Mathematics* contained notable contributions to the theory of MOLS. The third of these was the above-mentioned paper by Bose, Shrikhande, and Parker, showing the falsity of Euler's conjecture for all $n \equiv 2 \pmod 4$ other than $n = 2$ and 6. The fourth paper was also remarkable: S. Chowla, P. Erdős, and E. G. Straus showed that $N(n) \to \infty$ as $n \to \infty$. This disproved Euler's conjecture with a vengeance, and yet their proof was based only on the MacNeish results, a single recursive construction due to Bose and Shrikhande, and some number theory. Since then, much more has been revealed about the properties of $N(n)$ (see [22]).

A simple direct construction for a pair of orthogonal latin squares of order $3k + 1$ (such as 10 and 22) was given by Menon [23] in 1961: this was basically a reformulation of a construction in the Bose, Shrikhande, and Parker paper, but presented in very simple terms.

MOLS and projective planes

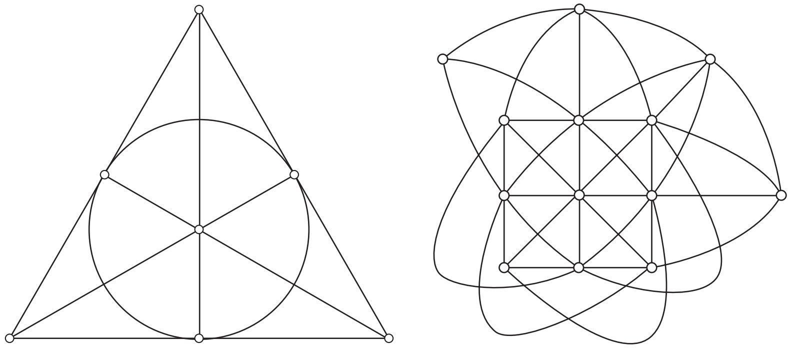
It is noteworthy that some of the results that could be used for the disproof of the Euler and MacNeish conjectures had already been available for a long time. In 1896 E. H. Moore [24] published a paper entitled 'Tactical memoranda I–III' which contained several results about MOLS, although they were not formulated in that language (see Chapter 10). They were subsequently rediscovered by later authors, unaware of their existence or the scope of their contents.

One such result was that

If n is a prime power, then there exists a set of $n - 1$ MOLS of order n;

this preceded MacNeish's publication of the same result by twenty-six years. The product construction used by MacNeish also appeared in Moore's paper. Curiously, there is a reference to Moore's 'Tactical memoranda' at the end of MacNeish's paper, as a place to find information about the application of Euler squares to contests between k teams of n members each – but there are no references to Moore in the part of the paper where MacNeish presents, in the language of Euler squares, the results also found in Moore's paper but with different terminology! Moreover, Moore's proofs and constructions were clearer than those of MacNeish, but for a long while it was MacNeish who was referred to in later work.

Such complete sets of MOLS (where $N(n)$ attains its maximum value of $n - 1$) are particularly important as they are equivalent to 'projective planes of order n'. A *finite projective plane* (see Chapter 10) is a geometry consisting of a finite number of points and lines, with the property that each pair of points lies on exactly one line, and each pair of lines meets in exactly one point (hence 'projective': even parallel lines meet, if projected to infinity); furthermore, to avoid trivial cases, there must exist at least four points, no three of which lie on a line.



The seven-point and thirteen-point geometries.

As we saw in Chapter 10, it can be shown that in a finite projective plane, each line contains the same number $s + 1$ of points; such a plane is said to be of *order s*. Using finite fields, one can prove that

A projective plane of order n exists whenever n is a prime power;

in this case, each line contains $n + 1$ points, each point lies on $n + 1$ lines, and the total numbers of points and lines are each $n^2 + n + 1$. Although Moore used finite fields in his construction of complete sets of MOLS, projective planes over finite fields had already been treated by K. G. C. von Staudt in 1856, and by 1850 Thomas Kirkman had proved the existence of projective planes of order n, when $n = 4$ or 8 or a prime number.

In 1936, in the second of two papers introducing balanced incomplete block designs, Frank Yates [25] explained how to obtain affine and projective planes from complete sets of MOLS; he referred to these as *completely orthogonalized latin squares* (hyper-Graeco-Latin squares). Yates noted that these exist for prime orders and for orders 4, 8, and 9, 'but higher non-primes have not been investigated'. Unaware of the results of Moore and MacNeish, Yates referred to R. A. Fisher's book *Design of Experiments* (the second edition from the same year) for this information. In later editions, Fisher noted that in 1939 W. L. Stevens had proved the existence for prime powers.

There is a record of Fisher's delight in this result. In a report in *Nature* in 1938 about a British Association meeting in Cambridge, with talks also by Norton, Youden, and Yates, Fisher wrote the following:

Mr. W. L. Stevens had a surprise in store, in the form of a demonstration of the fact that for any power of a prime a completely orthogonal square exists.

But it was a 1938 paper by Bose that was to prove the equivalence between projective planes of order n and sets of $n - 1$ MOLS of order n:

$$N(n) = n - 1 \textit{ if and only if there exists a projective plane of order } n.$$

Since projective planes of prime-power orders were well known, the result implied the existence of complete sets of MOLS of such orders. But Bose also gave a direct construction of complete sets of MOLS using finite fields: his proof of this result and his version of Moore and MacNeish's prime-power construction are quite similar to those used in most textbooks nowadays. Bose began his paper by stating that it was a conjecture of Fisher that a complete set of MOLS of order n exists for every prime power n, and he wrote that the constructions of affine and projective planes from such sets correspond exactly to those of Yates.

In 1949 H. Bruck and H. J. Ryser proved that if a projective plane of order n exists, where $n \equiv 1$ or $2 \pmod 4$, then n must be the sum of two perfect squares. This gives the non-existence of projective planes for $n = 14, 21, 22, \ldots$ (and for $n = 6$, but this was already known from Tarry's result). Their result was generalized the next year to give the celebrated *Bruck–Ryser–Chowla theorem* [26].

This connection between MOLS and projective planes makes it possible to relate a number of further results about projective planes, such as non-existence results, to MOLS. We note that, so far, no projective planes of non-prime-power order are known, so it could well be true that $N(n) = n - 1$ if and only if n is a prime power. After a long computer search based on coding-theoretic considerations, Lam [27] concluded that no projective plane of order 10 exists, so the value $n = 12$ is currently the smallest for which this statement is in doubt.

The influence from experimental design

In spite of Cretté de Palluel's 1788 application, little use was made of latin squares in experimental design until R. A. Fisher's boost to this area in the 1920s. Fisher was chief statistician at Rothamsted Experimental Station in Hertfordshire, UK, from 1919 to 1933, and later a professor in London, and wrote numerous papers on various aspects of statistics.

Fisher also conducted many experiments, often in agriculture, and took a great interest in the design of experiments, writing his classic book, *Design of Experiments*, in 1935. Many subjects from this area, including latin squares in field experiments, also appear in 1925 in his first book, and in 1926 he specifically wrote about the use of latin squares, and mutually orthogonal ones in particular, in experiments [28]. Moreover, Yates began his 1936 paper (mentioned earlier) with the observation

> Most biological workers are probably by now familiar with the methods of experimental design known as randomized blocks and the Latin square. These were originally developed by Prof. R. A. Fisher, when Chief Statistician at Rothamsted Experimental Station, for use in agricultural field trials.

In his *Design of Experiments*, Fisher mentioned earlier uses of latin squares in design. In particular, in a subsection of his latin square chapter on randomization, he warned against the use of systematic squares – that is, a preferred layout (latin square) used repeatedly. He displayed a specific square of order 5 with constant diagonals, and noted that if this square were used for elimination of soil differences in an agricultural experiment (as apparently it had been), it would fail as far as diagonal fertility differences were concerned.

Fisher also noted that this shortcoming had been realized by many others, and went on to mention the square below, 'known in Denmark since about 1872', as dealing with this difficulty.

$$\begin{matrix} A & B & C & D & E \\ C & D & E & A & B \\ E & A & B & C & D \\ B & C & D & E & A \\ D & E & A & B & C \end{matrix}$$

Although this square had been previously published in Denmark several times, it was credited to the Norwegian Knut Vik, who presented it in 1924.

Its apparent advantages notwithstanding, Fisher was not happy with the systematic use of this square, and while it is safe to say that a reference to *the* Knut Vik square is to the above square, there is confusion about the generalization of this terminology. All diagonals (including 'broken' diagonals) of the square are transversals, and squares with this property are now called *Knut Vik designs*; they exist if and only if n is not a multiple of 2 or 3. His square also happens to be an example of a *knight's move square*, because all the cells with the same symbol can be visited by a chess knight with allowable moves.

It seems, however, that the theory of latin squares has been more influenced by statisticians than by statistics. Certainly, some of the pioneers in the area of statistical design of experiments, although naturally interested in the practical uses of designs and in their enumeration, have also contributed significantly to the underlying theory. Some statisticians have actually warned against too much emphasis on latin squares in practice! Donald Preece has written that the prominent feature of latin squares in textbooks has 'led to their uncritical use', and even Fisher had an interesting remark:

> This experimental principle is best illustrated by the arrangement known as the Latin square, a method which is singularly reliable in giving precise comparisons,

adding

> when the number of treatments to be compared is from 4 to 8.

Still, in a standard work on statistical tables first published in 1938, Fisher and Yates provided tables of pairs of orthogonal Latin squares of orders 3 to 12 (except 6 and 10 – with a pair of order 10 included in later editions) as well as complete sets of MOLS of order 3, 4, 5, 7, 8, and 9, while on the topic of complete sets of MOLS Bose wrote that

> The work of Fisher and Yates has shown that such squares are of fundamental importance in experimental design.

Other results

Euler's 1782 paper left open several immediate challenges. In addition to his conjecture, he introduced such topics as enumeration and transversals. Naturally, new questions have come to light since then, and research on latin squares has both broadened and deepened. We continue this chapter by describing some of the more prominent recent developments.

Enumeration

A latin square of order n on symbols $1, 2, \ldots, n$ is *reduced* if both the first row and the first column are $1, 2, \ldots, n$ in order. A much-used enumeration function (see [29]) is $l(n)$, the number of reduced latin squares of order n; the total number of *distinct* latin squares on these symbols is then $n!(n-1)!\,l(n)$. Euler determined $l(n)$ for $n \leq 5$: the values are 1, 1, 1, 4, and 56.

A *latin rectangle* is a rectangular array of symbols in which no symbol occurs twice in any row or column. We let $l(k, n)$ denote the number of $k \times n$ latin rectangles on n symbols with first row $1, 2, \ldots, n$ and first column $1, 2, \ldots, k$. A result of Marshall Hall asserts that

Any latin rectangle of size $k \times n$ on n symbols can be completed to a latin square of order n.

Thus, it is interesting to examine $l(k, n)$, for all k, with a view to determining the specific value $l(n, n) = l(n)$.

For $k = 2$, Euler found the recursion

$$l(2, n) = (n-2)\,l(2, n-1) + (n-3)\,l(2, n-2).$$

He wrote down the first ten values of $l(2, n)$, but apparently did not recognize the problem of derangements here. However, Arthur Cayley made the connection in 1890, referring to the 'well-known problem' and stating the solution

$$l(2, n) = n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \cdots \pm \frac{1}{n!}\right) \Big/ (n-1).$$

In the same paper, Cayley recalculated Euler's values of $l(n)$, for small n.

In 1890 M. Frolov correctly stated that $l(6) = 9408$, but gave a value for $l(7)$ that was about thirteen times too large. In a series of papers from 1949–51, A. Sade showed that $l(7) = 16\,942\,080$, at the same time correcting earlier mistakes of S. M. Jacob and H. W. Norton. In 1969 M. B. Wells determined $l(8)$ to be 535 281 401 856. The value of $l(9)$ was found in 1975 by S. E. Bammel and J. Rothstein, and in 1995 B. D. McKay and E. Rogoyski determined $l(10)$; their paper also contains a listing of $l(k, n)$ for all $k \leq n \leq 10$. In 2005 $l(11)$ and the corresponding numbers of rectangles were found by McKay and I. M. Wanless; $l(11)$ has 48 digits. In 1981 Smetaniuk proved that the number of latin squares of a given order is strictly increasing, and the parameters determined in later years showed the possibilities and limitations of computers for all these enumeration problems.

Attempts at finding algebraic descriptions for latin square and rectangle enumeration have resulted in formulas that are impractical for calculation. Asymptotic results and bounds exist, however. In 1969, before the van der Waerden conjecture was proved (posed in 1926, this conjecture says that the permanent of an $n \times n$ doubly stochastic matrix $\mathbf{A}$ satisfies per$\mathbf{A} \geq n!/n^n$), Ryser [30] noted that its truth would imply that

$$l(n) \geq \frac{(n!)^{2n-2}}{n^{n^2}},$$

and in their 1992 textbook J. van Lint and R. M. Wilson proved that this bound is asymptotically of the right order of magnitude.

Transversals and partial transversals

As we have seen, transversals were important in Euler's 1782 paper – indeed, he considered them to be the first and foremost object of his paper, which contained the first results on latin squares without transversals. This line of investigation was continued in connection with research on orthogonality, since a latin square with no transversals cannot have an orthogonal mate. On the other hand, there are situations (for example, if the latin square is the multiplication table of a group) where the existence of just one transversal is enough to imply a decomposition into transversals, and thus the existence of an orthogonal mate. Euler had already exploited this.

In 1894 E. Maillet generalized Euler's findings that, within a certain class of latin squares, those of order $n = 2 \pmod 4$ have no transversals. A latin square of order mq is of *q-step type* if it can be obtained by replacing each individual entry in a cyclic latin square of order m by a latin square of order q, in such a way that these are on the same set of symbols if they replace the same symbol, and on disjoint sets of symbols otherwise. Maillet then proved that a latin square of order mq of q-step type has no transversals when m is even and q is odd; his proof is similar to the one given by Euler for $q = 1$, only more complicated.

While it has been known since Euler's time that there are latin squares with no transversals, two related problems have remained unanswered. In 1967 Ryser conjectured that every latin square of odd order has a transversal: this is true for latin squares coming from groups and for symmetric latin squares, where the main diagonal is a transversal. It has also been conjectured (by R. A. Brualdi, S. K. Stein, and probably others) that every latin square of order n has a *partial transversal* of length at least $n - 1$; that is, a set of cells in distinct rows, in distinct columns, and containing distinct symbols: this conjecture prompted people to search for long partial transversals. Since K. Koksma proved in 1969 that a latin square of order $n \geq 7$ has a transversal of length at least $\frac{1}{3}(2n + 1)$, there have been a number of improvements; for example, in 2008 Hatami and Shor [31] proved the existence of a partial transversal of length at least $n - 11.1(\log n)^2$.

Quasigroups, completion, and critical sets

The Cayley table (or multiplication table) of any finite group is a latin square, so when group theory began to flourish in the second half of the 19th century there were plenty of latin squares to study. It seems, however, that few of the leading group-theorists were interested in this aspect, Cayley himself being one of them.

Not every latin square bordered by (say) its first row and column is the Cayley table of a group, since associativity may not hold. Instead, it corresponds exactly to the concept of a *quasigroup*, an algebraic structure with a binary operation for which the equation $x \cdot y = z$ has unique solutions for x (given y and z) and for y (given x and z). Quasigroups were essentially introduced by E. Schröder in the 1870s, but did not achieve general attention until the 1930s. Since then, many results on latin squares have been first formulated in terms of quasigroups.

One particular area that has often been discussed in terms of quasigroups is that of the completion of partial structures. In terms of latin squares, the typical question is whether a *partial latin square* (a square array, possibly with empty cells, whose entries do not violate the latin square conditions) can be completed to a latin square. A conjecture that was posed in 1960, known as the *Evans conjecture*, was proved true by Smetaniuk in 1981. It states that a partial latin square of order n with at most $n - 1$ non-empty cells can always be completed. In quasigroup terminology, such problems are often formulated with extra conditions, such as idempotency, commutativity, etc.

A partial latin square is a *critical set* if it can be completed in exactly one way to a latin square of the same order, and if the deletion of any entry destroys this uniqueness. The problem of sizes of critical sets is a topic with many open questions, and may recently have become especially popular because of its connections with sudoku puzzles.

Sudoku

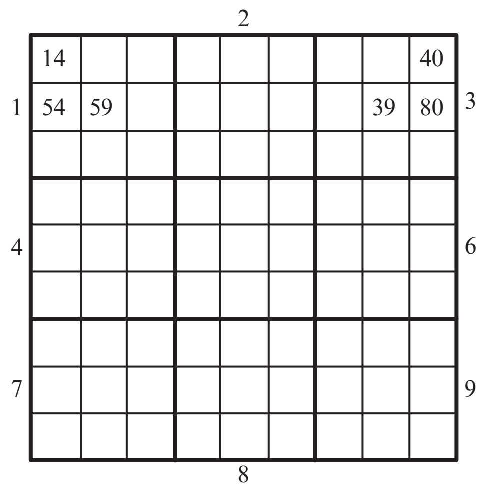
Since around 2005, sudoku puzzles have been challenging crosswords as the favourite pastime offered by newspapers and magazines all over the world, and multiple collections of sudokus have been issued in many countries. A completed *sudoku* is a latin square of order 9 with symbols $1, 2, \ldots, 9$, with the additional condition that each of the nine natural 3×3 subsquares contains all nine symbols. The task is to complete the square from a given set of prescribed entries; this set must be a *defining set* for the sudoku – that is, there is a unique solution, and it is often required that the set be *critical*, so that for any proper subset of the 'givens' there is more than one solution. It has recently been shown that there is no such defining set of size 16 (see [32]) – in other words, a sudoku puzzle such as the following, with a set of 17 givens, represents the least possible number of givens for a unique solution:

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

A sudoku puzzle with the minimum number of givens.

Sudokus have their name from Japanese, but they were invented in the United States in the late 1970s when the architect Howard Garns began publishing such puzzles in Dell magazines, the first appearing in *Dell Pencil Puzzles and Word Games* in May 1979. The puzzles were then called 'number place', but when they became popular in Japan in the mid-1980s, it was under the name *suji wa dokushin ni kagiru*, which means ‘number is limited only single’; this became abbreviated to ‘sudoku’. In 1986, when the interest exploded, the Japanese publishing company Nikoli introduced the extra rule that the cells filled with prescribed entries should have 180-degree rotational symmetry, though this rule is now often ignored. It is believed that the least possible number of givens for a sudoku puzzle with 180-degree symmetry is 18, but this has not yet been proved.

Sudokus have often been attributed to Euler, but this is incorrect. However, as has been pointed out by Christian Boyer [33], there were forerunners of sudokus in France as early as the end of the 19th century, mainly in the form of puzzles about completing magic squares. The following figure shows one such from 1891, where subsquares of sudoku type were part of the problem: complete the magic square with numbers 1–81 so that, in each of the indicated subsquares, all row and column sums are 123.



A sudoku-like problem about magic squares from 1891.

The following figure shows a square from a problem from 1895: complete the square with numbers 1–9 so that the sum of the entries in each row, column, and main diagonal is the same. The solution actually turns out to comply with the sudoku rules! The solution, published in 1895, is also shown.

7	8	9	1	2	3	4	5	6
3				4				8
5				9				1
8				3				4
1	2	3	4	5	6	7	8	9
6				7				2
9				1				5
2				6				7
4	5	6	7	8	9	1	2	3

7	8	9	1	2	3	4	5	6
3	1	2	6	4	5	9	7	8
5	6	4	8	9	7	2	3	1
8	9	7	2	3	1	5	6	4
1	2	3	4	5	6	7	8	9
6	4	5	9	7	8	3	1	2
9	7	8	3	1	2	3	4	5
2	3	1	5	6	4	8	9	7
4	5	8	7	6	9	1	2	3

A problem from 1895 and its published sudoku solution.

Sudoku squares are also included in W. T. Federer's 1955 book [34] on experimental designs, in the general form of latin squares of order $n = m_1 m_2$ which satisfy the requirement that the entries in each natural $m_1 \times m_2$ subsquare are all distinct. Federer attributed this to G. M. Cox, and he called such a latin square *magic* ('super magic' if $m_1 \neq m_2$ and the $m_2 \times m_1$ subsquares also have the property). This concept makes better sense than just the row and column requirements do in (say) an agricultural experiment designed to expose the treatments more evenly to varying conditions in the field. In 1956 W. U. Behrens [35] published an extension of this idea to *gerechte designs* (proper designs), where the subsquare pattern is replaced by any given partition of the latin square into n regions of n cells each.

Wayne Gould, a retired judge and a resident of Hong Kong, realized the power of computer programs in generating sudokus, and on 12 November 2004 *The Times* of London published one of his sudokus (or rather, a Su Doku, as it was then called). Many more newspapers followed as the puzzles became increasingly popular, and *The Times* had Gould answer readers' questions on sudokus; he later produced sudokus for several newspapers. In the US, sudoku championships, live TV sudoku shows, and similar events followed, as a further sign of their popularity.

It has been calculated that there are 6 670 903 752 021 072 936 960 distinct sudoku squares. This means that only about one in a million 9×9 latin squares is a sudoku square. Many of these can be obtained by performing simple operations on others, and the total number of *essentially different* sudoku squares is 5 472 730 538. It is hard to compare this to a number for latin squares, as the two have very different symmetries (see [36]).

There are now several variants and generalizations of sudoku, many of these also interesting to mathematicians. Other grid sizes than the usual 9×9 are possible, with other subsquare sizes. The grid can also be of more than two dimensions. Further, subdividing the basic square – which still must be a latin square when filled in – into shapes that are not necessarily similar is a possibility (corresponding to the gerechte designs mentioned above). Also, rather than just requiring the entries in a subarea to be distinct, one can have conditions on their sum, product, or the like. An example of the latter variants is KenKen, where no entries are prescribed; the difficulties and uniqueness come from the shapes (called cages) and requirements alone. But the solution is always a good old latin square!

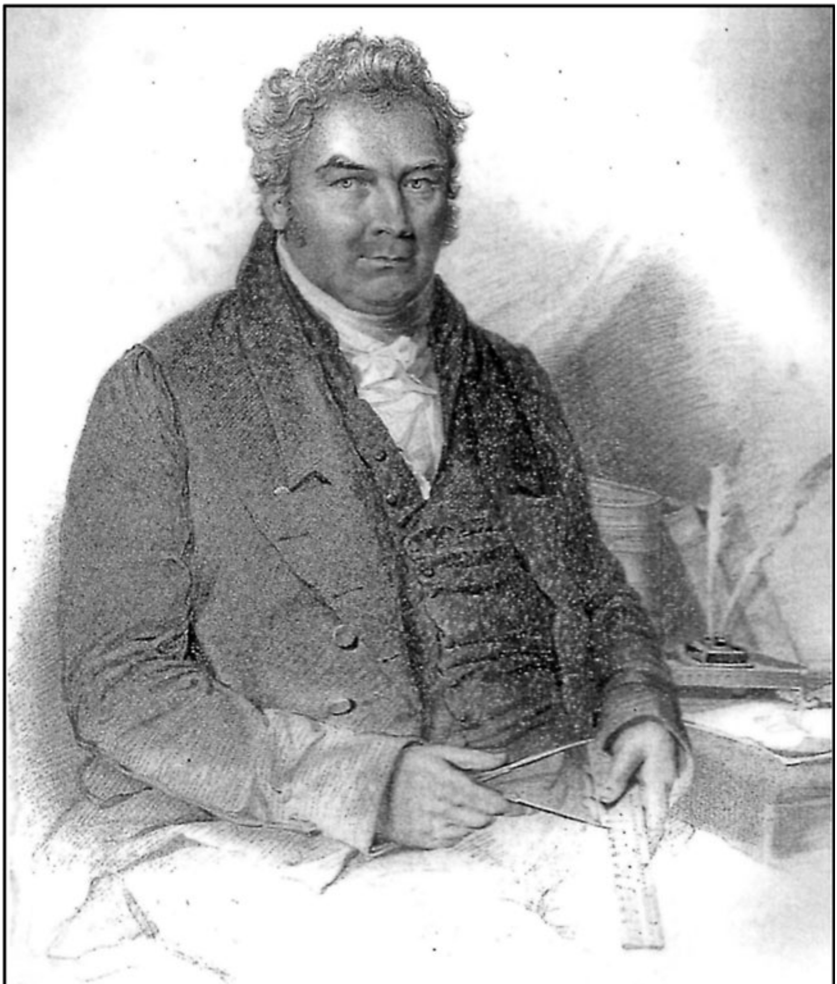
Notes and references

1. The illustration appears in D. B. MacDonald, Description of a silver amulet, *Z. Assyriologie und Verwandte Gebiete* 26 (1912), 267–9, and also in S. Seligmann, Das Siebenschläfer-Amulett, *Der Islam* 5 (1914), 370–88. Note that the penultimate entry in the last column is incorrect.
2. The illustration and the instruction are taken from J. Shurreef, *Qanoon-E-Islam* (1832), 'composed under the direction of', and translated into English by, G. A. Herklots as *Customs of the Moosulmans of India; Comprising a Full and Exact Account of Their Various Rites and Ceremonies from the Moment of Birth Till the Hour of Death*. The same latin square (except that the last three rows are permuted) also appears in the book of al-Būnī mentioned below (with Arab numbers) and in both versions in E. Doutté, *Magie et Religion dans l'Afrique du Nord* (1908), which also gives detailed instructions on how to use this square against infidelity. Doutté's book contains many such examples of early latin squares.
3. This book has been reissued in Arab countries, but there are no translations into any other language. As can be seen, many illustrations are quite defective – a fact that we can use when we try to investigate the constructions behind the squares presented.
4. This is argued in W. Ahrens, Die "magischen Quadrate" al-Būnī's, *Der Islam* 12 (1922), 157–77.
5. N. Pandit, *Ganita-kaumudi* (1356) (republished in The Princess of Wales Saravati Bhavana Sanskrit Series, No. 57, Part II).

6. The card problem is sometimes called ‘Bachet’s square’ from its appearance in C. G. Bachet de Méziriac’s *Problèmes Plaisans et Délectables qui se font par les Nombres*, which first appeared in 1612 but underwent several changes in later editions; in fact, the card puzzle may first have appeared as late as the 1884 revision. It featured in the 1723 edition of J. Ozanam’s *Récréations Mathématiques et Physiques*, where the solutions were enumerated (wrongly). For a discussion of its appearance in various editions of these books, see P. Ulrich, Officers, playing cards, and sheep, *Metrika* 56 (2002), 189–204. Our figure shows one solution from a later edition of Bachet, and a different solution from the front page of a modern textbook.
7. F. C. de Palluel, Sur les avantages et l’economie que procurent les racines employées à l’engrais des moutons à l’étable, *Mémoires d’Agriculture, Trimestre d’Eté* (1788), 17–23 (English transl., On the advantage and economy of feeding sheep in the house with roots, *Ann. Agriculture* 14 (1790), 51–5).
8. A nice collection of stamps arranged in a latin square appears on-line in *Some comments on* 4 × 4 *philatelic Latin Squares*, by P. D. Loly and G. P. H. Styan at: http://chance.amstat.org/wp-content/themes/mimbo/supplemental/styan.pdf.
9. This appears already in Pandit (note 5). Here it is taken from S. Cammann, Islamic and Indian magic squares. Part II, *Hist. Religions* 8 (1969), 271–99.
10. Further to note 4, another paper by W. Ahrens, Studien über die “magischen Quadrate” der Araber, *Der Islam* 7 (1917), 186–250, mentions other early works in which orthogonal latin squares occur. See also D. Knuth’s homepage http://www-cs-faculty.stanford.edu/∼ uno/orthogonal.txt for a brief discussion.
11. J. Sauveur, Construction générale des quarrés magiques, *Mém. Acad. Royales des Sciences* (1710), 92–138.
12. In the book *Traité des Quarrés Sublimes* (1704), Poignard used the term ‘Quarré du Soleil par Progression Arithmétique Répétée’ for a latin square.
13. The treatise *Ku su ryak* (Summary of the Nine Branches of Numbers) by Choe Sŏk-chŏng was described by H. Akira in 1983, Magic squares on “Ku su Ryak” by Chai Suk-Jong, *J. Hist. Math., Japan* 99, 1–2 [in Japanese], by Hong-Jeop Song in a presentation at the 2008 Global KMS Conference in Jeju, Korea, and by Ko-Wei Lih in the paper, A remarkable Euler square before Euler, *Math. Mag.* 83 (2010), 163–7.
14. G. Tarry, Le problème des 36 officiers, *Comptes Rendus* 29 (1900), 170–203.
15. Schumacher’s letter is discussed in H. Gropp, “Gaußsche Quadrate” or Knut Vik designs – the history of a combinatorial structure, *Proc. 2nd Gauss Symposium. Conference A: Mathematics and Theoretical Physics, Munich, August 2–7 1993* (ed. M. Behara, R. Fritsch, and R. G. Lintz), de Gruyter, Symposia Gaussiana (1995), 121–34.
16. R. A. Fisher and F. Yates, The 6 × 6 latin squares, *Proc. Cambridge Philos. Soc.* 30 (1934), 492–507.
17. D. Betten, Zum Satz von Euler-Tarry, *Math. Naturwiss. Unterricht* 36 (1983), 449–53.

18. D. R. Stinson, A short proof of the nonexistence of a pair of orthogonal squares of order six, *J. Combin. Theory (A)* 36 (1984), 373–6.
19. H. F. MacNeish, Euler squares, *Ann. Math.* 23 (1922), 221–7.
20. R. C. Bose, S. S. Shrikhande, and E. T. Parker, Further results on the construction of mutually orthogonal latin squares and the falsity of Euler's conjecture, *Canad. J. Math.* 12 (1960), 189–203.
21. M. Gardner, Euler's spoilers: The discovery of an order-10 Graeco-Latin square, *Martin Gardner's New Mathematical Diversions from Scientific American*, Allen and Unwin (1966), 162–72.
22. S. Chowla, P. Erdős, and E. G. Straus, On the maximal number of pairwise orthogonal latin squares of a given order, *Canad. J. Math.* 12 (1960), 204–8; this paper actually proved that $N(n) > \frac{1}{3}n^{1/91}$, for large n. In 1973, Rick Wilson published the bound $N(n) > n^{1/17} - 2$, for large n. As in the 1960 paper, number theory plays a big part, but whereas Chowla, Erdős, and Straus could write 'Our proof involves no new combinatorial insights', Wilson's paper contains the first form of his important theorem on constructions of transversal designs. Later improvements on this bound include one by Lu, who in 1985 proved that $N(n) \geq n^{10/143} - 2$. Upper bounds are rarer, but it is known that arbitrarily large gaps between $N(n)$ and $n - 1$ can occur.
23. P. K. Menon, Method of constructing two mutually orthogonal latin squares of order $3n + 1$, *Sankhya* A XXIII (1961), 281–2.
24. E. H. Moore, Tactical memoranda I–III, *Amer. J. Math.* 18 (1896), 264–303.
25. F. Yates, Incomplete randomized blocks, *Ann. Eugenics* 7 (1936), 121–40.
26. H. Bruck and H. J. Ryser, The non-existence of certain finite projective planes, *Canad. J. Math.* 1 (1949), 88–93, and S. Chowla and H. J. Ryser, Combinatorial problems, *Canad. J. Math.* 2 (1950), 93–9.
27. C. W. H. Lam, The search for a finite projective plane of order 10, *Amer. Math. Monthly* 98 (1991), 305–18.
28. R. A. Fisher, The arrangement of field experiments, *J. Ministr. Agric.* 33 (1926), 503–13.
29. Some other enumeration problems are to establish the number of isomorphism classes, the number of isotopy classes, and the number of main class isotopy classes of latin squares of order n. The seventeen types of latin square of order 6, found by Tarry in his solution of the officers problem, can be identified as coming from the twenty-two isotopy classes for order 6 by merging ten of them into five pairs.
30. H. J. Ryser, Permanents and systems of distinct representatives, *Combinatorial Mathematics and Its Applications* (ed. R. C. Bose and T. A. Dowling), University of Carolina Press (1969), 55–70.
31. P. Hatami and P. W. Shor, A lower bound for the length of a partial transversal in a latin square, *J. Combin. Theory (A)* 115 (2008), 1103–13.
32. G. McGuire, B. Tugemann, and G. Civario, There is no 16-clue sudoku: Solving the sudoku minimum number of clues problem, arXiv:1201.0749v1 (2012).

33. C. Boyer, Sudoku's French ancestors, *Math. Intelligencer* 29(1), 37–44; and solutions to some problems appeared in 29(2), 59–63.
34. W. T. Federer, *Experimental Design: Theory and Applications*, Macmillan (1955).
35. W. U. Behrens, Feldversuchsanordnungen mit verbessertem Ausgleich der Bodenunterschiede, *Z. Landwirtschaftliches Versuchs- und Untersuchungswesen* 2 (1956), 176–93. See also R. A. Bailey, P. J. Cameron, and R. Connelly, Sudoku, gerechte designs, resolutions, affine space, spreads, reguli, and Hamming codes, *Amer. Math. Monthly* 115 (2008), 383–404.
36. See http://www.afjarvis.staff.shef.ac.uk/sudoku/sudgroup.html for the calculations and an explanation of what is meant by 'essentially different'.



Peter Nicholson (1765–1844), practical builder and mathematician.

CHAPTER 12

Enumeration (18th–20th centuries)

E. KEITH LLOYD

By the end of the 18th century, what is now called 'enumerative combinatorics' was emerging as a distinct discipline. The connections between certain combinatorial problems and algebraical expansions had already been recognized, but were now more extensively exploited. In the 20th century the theory of permutation groups was successfully used to solve many enumeration problems.

Introduction

In Chapters 5 and 6 Eberhard Knobloch discusses the development of permutations, combinations, etc., up to the mid 18th century. He mentions in particular that in the 17th century, authors often inserted sections on combinatorics into their textbooks on arithmetic and algebra, or began to write special monographs on the subject. But he also points out that some of the contributors, such as Abraham de Moivre, were more interested in combinatorics applied to games than as a subject in its own right.

By the end of the 18th century, combinatorial analysis (as it was then usually called) was recognized as a distinct discipline, and rather more substantial books began to appear devoted exclusively to the topic. It had already been realized that there is an intimate connection between certain combinatorial problems and algebraical expansions, and so in the books then appearing the authors spent much time seeking algorithms for calculating coefficients in various algebraical expressions, or for rewriting algebraical expressions in different forms.

In the 19th century, group theory was developing as a subject, and by the end of the century some elementary use of it had been made in combinatorics. In the first half of the 20th century, deeper interrelationships between permutation group theory and combinatorics were recognized, notably by J. Howard Redfield and George Pólya. Unfortunately, Redfield's work was overlooked and remained unpublished until many years after his death; consequently, many of his results were rediscovered independently by other researchers later in the 20th century.

Enumeration and algebraical expansions

The starting point for this topic is the fact that combinations are intimately connected with expansions of binomial factors. The observation that there is such a connection was attributed to Thomas Harriot (1560–1621) by Peter Nicholson, who in 1818 wrote [34, pp. v–vi]:

> In the case of products formed of binomial factors, of the form $x + a, x + b, x + c$, &c., it had, long before my time, been observed by Hariot, that the coefficients of the second term were the sum of all the quantities a, b, c, &c.; the coefficients of the third term, the sum of all the parcels of the second order of combination; the coefficient of the fourth term, the sum of all the parcels in the third order; and so on.

In modern terminology, this can be explained as follows: if the set $S = \{a, b, c\}$, then all the *subsets* (also called *combinations without repetition*) of S are 'generated' by the expansion

$$(x + a)(x + b)(x + c) = x^3 + (a + b + c)x^2 + (bc + ac + ab)x + abc.$$

Here the summands in the coefficient of x^2 (Nicholson's second term) correspond to the 1-subsets (subsets of size 1) of S, the summands in the coefficient of x (his third term) correspond to the 2-subsets of S, and the constant term (his fourth term) corresponds to the single 3-subset of S. For completeness, one must adopt the convention that the coefficient of x^3 corresponds to the 0-subset (or empty subset).

The mismatch between the powers of x and the sizes of the subsets can be avoided by considering the equivalent expansion

$$(1 + ax)(1 + bx)(1 + cx) = 1 + (a + b + c)x + (bc + ac + ab)x^2 + (abc)x^3,$$

where the summands in the coefficient of x^r correspond to the r-subsets of S.

For an n-set $S_n = \{a_1, a_2, \ldots, a_n\}$, the subsets are generated in the same way by the expansion of

$$(x + a_1)(x + a_2) \ldots (x + a_n),$$

or by

$$(1 + a_1x)(1 + a_2x) \ldots (1 + a_nx).$$

To *count* subsets of S, rather than to *generate* them, we put $a = b = c = 1$ to get

$$(1 + x)^3 = 1 + 3x + 3x^2 + x^3,$$

or, more generally,

$$(1 + x)^n = 1 + C(n, 1)x + \cdots + C(n, r)x^r + \cdots + C(n, n)x^n,$$

where the binomial coefficient $C(n, r)$ (for $r = 0, 1, \ldots, n$) is equal to the number of r-subsets of an n-set – that is, the number of ways of choosing r objects from n objects (the number of *r-combinations of n objects*). The history of this interpretation of binomial coefficients is discussed in Chapters 4–7.

The above ideas may be extended to include repetition of elements. For example, the coefficient of x^r in the expansion of

$$(1 + ax + a^2x^2)(1 + bx)(1 + cx)$$
$$= 1+(a + b + c)x+(a^2 + bc + ac + ab)x^2+(a^2b + a^2c + abc)x^3+(a^2bc)x^4$$

generates the r-combinations of a, b, c in which a may be included up to two times, but b and c cannot be repeated. If a may be repeated an unlimited number of times, then the factor involving a is

$$(1 + ax + a^2x^2 + a^3x^3 + \cdots) = (1 - ax)^{-1}.$$

Hence, r-combinations of $a_1, a_2, \ldots, a_n$ with unrestricted repetition are generated by

$$\Big((1 - a_1x)(1 - a_2x) \ldots (1 - a_nx)\Big)^{-1}.$$

As above, the combinations are counted by putting $a_1 = a_2 = \cdots = a_n = 1$, to give

$$(1 - x)^{-n}$$

as the generating function for combinations with repetition – that is, the coefficient of x^r in the expansion of $(1 - x)^{-n}$ is the number of *r-combinations with repetition from an n-set*.

The theory of partitions (see Chapter 9) is another area in which generating functions have been used, notably by Leonhard Euler. For example, the coefficient of $x^r z^s$ in the expansion of

$$\big((1 - xz)(1 - x^2 z)(1 - x^3 z)(1 - x^4 z)\ldots\big)^{-1}$$

is the number $p_s(r)$ of partitions of r into s parts.

From the above, it can be seen that an important part of enumerative combinatorics is to develop efficient methods for expanding products and powers of polynomials and power series; this was enthusiastically taken up by the Hindenburg School of Combinatorics, led by Carl Friedrich Hindenburg (1741–1808), professor of physics at Leipzig. Haas [16] writes that

> Hindenburg and his school attempted, through systematic development of combinatorials, to give it a key position within the various mathematical disciplines. Combinatorial considerations, especially appropriate symbols, were useful in the calculations of probabilities, in the development of series, in the inversion of series, and in the development of formulas for higher differentials.

Haas also stated that a 'central problem' of Hindenburg was to find an expression for the b_i, explicitly in terms of the a_i, in the expansion

$$(a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m)^n = b_0 + b_1 x + b_2 x^2 + \cdots + b_{mn} x^{mn}.$$

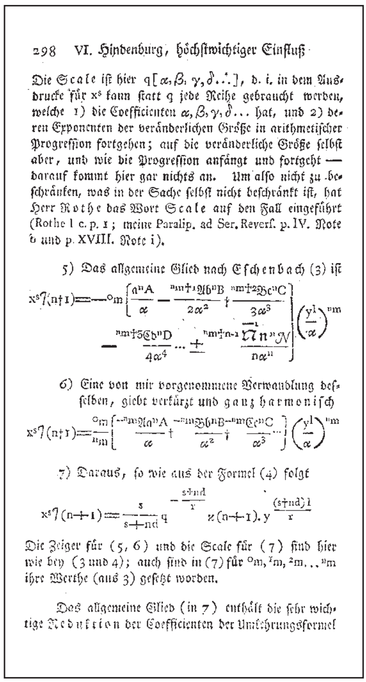
The importance which Hindenburg attached to such expansions can be judged by the title which he gave to a book [21] published in 1796: *Der polynomische Lehrsatz, das wichtigste Theorem der ganzen Analysis nebst einigen verwandten und andern Sätzen* (The Polynomial Theorem, the Most Important Theorem in the Whole of Analysis, together with Several Related and Other Theorems). Hindenburg wrote about half of the book, but, as indicated on the title page, there were also contributions from other members of the school (Tetens, Klügel, Kramp, and Pfaff), to which Hindenburg added many footnotes.

The school had little, if any, influence outside Germany; a list of most of the writings of the school is given by Weingärtner [49]. A feature commented upon by various later writers is the complicated notation used by the school. For example, even though Nicholson spoke favourably of some of Hindenburg's work, he also wrote [34, p. xxiv]:

> Indeed, Hindenburgh's notation was too complex to be introduced into an English publication, as he uses no less than six or seven different kinds of letters, taken from the Grecian, Roman, and Gothic alphabets; and also different sizes of letters, both sloping and upright, which render it extremely perplexing.

A sample page from the book is shown below. It must have been a problem for the typesetter; even the addition signs are in assorted fonts and sizes. Another writer once said that Hindenburg disguised everything under a thick layer of notation.

Der
polynomische Lehrsatz
das
wichtigste Theorem
der
ganzen Analysis
nebst
einigen verwandten und andern Sätzen
Neu bearbeitet und dargestellt
von
Tetens, Klügel, Kramp, Pfaff und Hindenburg.
Zum Druck befördert
und mit
Anmerkungen, auch einem kurzen Abrisse der combinatorischen Methode und ihrer Anwendung auf die Analysis
versehen
von
Carl Friedrich Hindenburg.
Leipzig
bei Gerhard Fleischer dem Jüngern
1796.



298 VI. Hindenburg, höchstwichtiger Einfluß

Die Scale ist hier $q[\alpha, \beta, \gamma, \delta \ldots]$, d. i. in dem Ausdrucke für x^s kann statt q jede Reihe gebraucht werden, welche 1) die Coefficienten $\alpha, \beta, \gamma, \delta \ldots$ hat, und 2) deren Exponenten der veränderlichen Größe in arithmetischer Progression fortgehen; auf die veränderliche Größe selbst aber, und wie die Progression anfängt und fortgeht — darauf kommt hier gar nichts an. Um also nicht zu beschränken, was in der Sache selbst nicht beschränkt ist, hat Herr Rothe das Wort Scale auf den Fall eingeführt (Rothe l c. p. 1; meine Paralip. ad Ser. Reverf. p. IV. Note b und p. XVIII. Note i).

5) Das allgemeine Glied nach Eschenbach (3) ist

$$x^s 7(n+1) = {}^0m \left\{ \frac{a^n A}{\alpha} - \frac{{}^{nm+1}\mathfrak{A} b^n B}{2\alpha^2} + \frac{{}^{nm+2}\mathfrak{B} c^n C}{3\alpha^3} - \frac{{}^{nm+3}\mathfrak{C} d^n D}{4\alpha^4} \ldots \pm \frac{{}^{nm+n-1}\overline{\mathfrak{N}}\, n^n \mathcal{N}}{n\alpha^n} \right\} \left(\frac{y^l}{\alpha}\right)^{nm}$$

6) Eine von mir vorgenommene Verwandlung desselben, giebt verkürzt und ganz harmonisch

$$x^s 7(n+1) = \frac{{}^0m}{{}^nm} \left\{ \frac{{}^{-nm}\mathfrak{A} a^n A}{\alpha} + \frac{{}^{-nm}\mathfrak{B} b^n B}{\alpha^2} + \frac{{}^{-nm}\mathfrak{C} c^n C}{\alpha^3} \ldots \right\} \left(\frac{y^l}{\alpha}\right)^{nm}$$

7) Daraus, so wie aus der Formel (4) folgt

$$x^s 7(n+1) = \frac{s}{s+nd}\, q^{-\frac{s+nd}{r}}\, \varkappa(n+1).\, y^{\frac{(s+nd)l}{r}}$$

Die Zeiger für (5, 6) und die Scale für (7) sind hier wie bey (3 und 4); auch sind in (7) für ${}^0m, {}^1m, {}^2m, \ldots {}^nm$ ihre Werthe (aus 3) gesetzt worden.

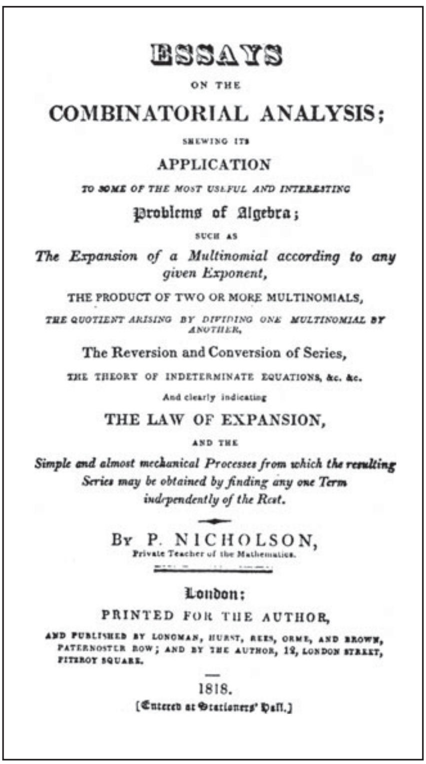
Das allgemeine Glied (in 7) enthält die sehr wichtige Reduktion der Coefficienten der Umkehrungsformel

Title page and a sample page from *Der polynomische Lehrsatz.*

Peter Nicholson, a Scotsman, was largely self-taught. An architect by profession, he wrote about twenty-five books, mainly on carpentry, architecture, and building, but most of them have a strong mathematical bias. Between 1817 and 1824 he published several purely mathematical books, in some of which he is described as a 'Private Teacher of the Mathematics'; his *Essays on the Combinatorial Analysis* [34] appeared in 1818. From the thirty-two-page introduction it appears that he originally planned two essays (II and III in the completed work), but he had a tendency to rewrite his books when they were in press. He decided to prefix his first essay as an introduction to the second, and at a very late stage he 'found it necessary . . . to add a Fourth Essay . . .'. It was just as well that he did, for a note on the back of the title page to the complete work reads:

> The Reader is desired to peruse pages 49, 50, and 51, of the Fourth Essay, which explain the Notation employed in the Introduction and in the Essays themselves, before he begins to read any part of the Work.

The work appears to be the first book in English with the words 'combinatorial analysis' in the title, but certainly it would have benefited from being rewritten and reorganized.



ESSAYS
ON THE
COMBINATORIAL ANALYSIS;
SHEWING ITS
APPLICATION
TO SOME OF THE MOST USEFUL AND INTERESTING
Problems of Algebra;
SUCH AS
The Expansion of a Multinomial according to any given Exponent,
THE PRODUCT OF TWO OR MORE MULTINOMIALS,
THE QUOTIENT ARISING BY DIVIDING ONE MULTINOMIAL BY ANOTHER,
The Reversion and Conversion of Series,
THE THEORY OF INDETERMINATE EQUATIONS, &c. &c.
And clearly indicating
THE LAW OF EXPANSION,
AND THE
Simple and almost mechanical Processes from which the resulting Series may be obtained by finding any one Term independently of the Rest.

BY P. NICHOLSON,
Private Teacher of the Mathematics.

London:
PRINTED FOR THE AUTHOR,
AND PUBLISHED BY LONGMAN, HURST, REES, ORME, AND BROWN, PATERNOSTER ROW; AND BY THE AUTHOR, 12, LONDON STREET, FITZROY SQUARE.
1818.
[Entered at Stationers' Hall.]

Title page of Nicholson's *Essays on the Combinatorial Analysis*.

The introduction to Nicholson's *Essays* is valuable from a historical point of view, since he discussed what other authors he had studied, and at what stage in his own researches he had become aware of them. But he was rather chauvinistic and, after writing

> . . . in this country, where genius abounds, and where the combinatorial analysis began to dawn . . . ,

he then credited Abraham de Moivre [9], [10]:

> Thus it appears, that M. Ab. Demoivre was the inventor of the Combinatorial Analysis. To foreigners alone we owe the subsequent improvements and advancement of this branch of mathematical science.

It seems that Nicholson's motivation for taking up the subject came from studying William Emerson's *Increments* [12], and Nicholson himself wrote a book [33] on that subject. Part way through his work on combinatorics, Nicholson became aware of what he termed 'Hindenburgh's Combinatorial Essays, in 2 vols. Leipsic, 1796'; presumably he was referring to [21] and [22]. Nicholson's friend J. A. Hamilton (described as a professor of music) translated parts of Hindenburg's work for him.

Nicholson's *Essay I* is entitled *Combinatorial Analysis. General Principles of Combinations and Permutations*, but it also contains material on partitions of numbers, which he called *decompositions*. Much of the material consists of systematically listing or counting combinations, etc., and an appendix (on partitions) to *Essay I* was written by Hamilton. *Essay II*, entitled *Combinatorial Analysis Applied to Series in General*, includes expansions for binomials and multinomials, and for quotients of series. Also considered is the problem of reverting series:

Given y as a series in x, find x as a series in y.

The first part of *Essay III*, entitled *Principles of Binomial Factors . . .* , is concerned with expressing products of binomial factors as sums of products of other binomial factors. A typical problem is:

Given two sets of numbers a, b, c, d, and $\alpha, \beta, \gamma, \delta$, find A, B, C, D, and E such that

$$\begin{aligned}&(x+a)(x+b)(x+c)(x+d)\\ &\quad = A(x+\alpha)(x+\beta)(x+\gamma)(x+\delta) + B(x+\alpha)(x+\beta)(x+\gamma)\\ &\qquad + C(x+\alpha)(x+\beta) + D(x+\alpha) + E. \qquad (*)\end{aligned}$$

Nicholson gave a quite efficient algorithm for this (which works for n factors, not just for four). When the numbers a, b, c, d are in arithmetic progression, Nicholson called the product $(x+a)(x+b)(x+c)(x+d)$ a *factorial*. Partway through [33], he introduced the following notation for such factorials:

$$(x+r)^m|^d = (x+r) \times (x+r+d) \times (x+r+2d) \times \cdots \times \big(x+r+(m-1)d\big);$$

for example,

$$x^5|^2 = x(x+2)(x+4)(x+6)(x+8).$$

In [34], Nicholson modified this notation by shortening the vertical bar, to $(x+r)^{m|d}$; in addition, if $d = -r$ was negative he wrote $\bar{r}$ rather than $-r$, giving, for example,

$$(x+2)^{4|\bar{3}} = (x+2)(x-1)(x-4)(x-7).$$

A special case of interest (not mentioned by Nicholson) is when $a = b = c = d = 0$ and $\alpha = 0, \beta = -1, \gamma = -2, \delta = -3$; then $(*)$ becomes

$$x^4 = Ax(x-1)(x-2)(x-3) + Bx(x-1)(x-2) + Cx(x-1) + Dx + E,$$

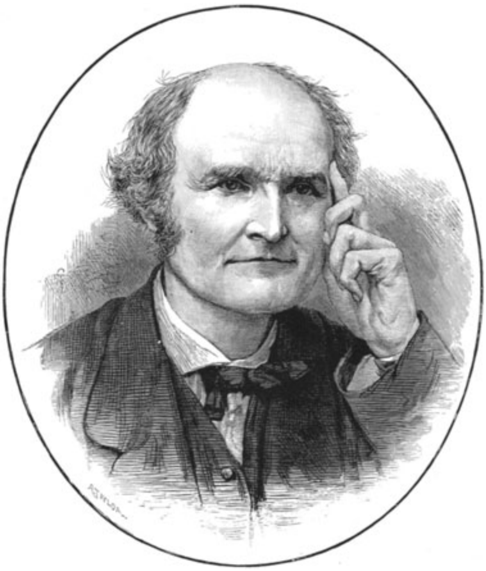
where A, B, C, D, E are the *Stirling numbers of the first kind*. Similarly, when $a = 0$, $b = -1$, $c = -2$, $d = -3$, and $\alpha = \beta = \gamma = \delta = 0$, then A, B, C, D, E are the *Stirling numbers of the second kind*. Later in *Essay III* Nicholson studied similar expansions in which the factorials are in the denominators; in these cases, there are infinitely many terms on the right-hand sides of the expansions.

In some ways, apart from the explanation of the notation, *Essay IV* (on *Figurate Numbers*) is redundant. The nth order of figurate numbers are just the numbers in the nth row of *Pascal's triangle* when it is laid out in the following rectangular form:

$$\begin{array}{cccccc} 1 & 1 & 1 & 1 & 1 & \cdots \\ 1 & 2 & 3 & 4 & 5 & \cdots \\ 1 & 3 & 6 & 10 & 15 & \cdots \\ 1 & 4 & 10 & 20 & 35 & \cdots \\ 1 & 5 & 15 & 35 & 70 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{array}.$$

The work of Cayley and Jordan on trees

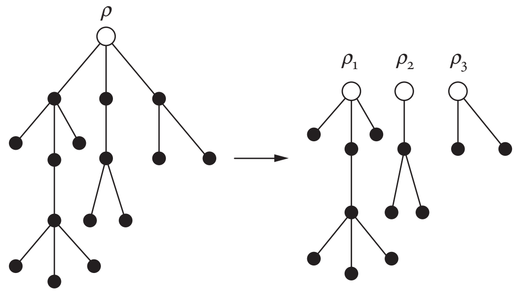
In the mid 19th century, the term *tree* was introduced in a mathematical context by Arthur Cayley, and various types of trees were enumerated (see Chapter 8). Around the same time, chemists were beginning to clarify their ideas about valency and the structure of molecules, and the work on trees soon proved to be relevant for counting the numbers of isomers of various chemical compounds.



Arthur Cayley (1821–95).

As we saw in Chapter 8, a *tree* is a structure consisting of a set of vertices, certain pairs of which are joined by edges, subject to the restriction that the structure is connected and contains no cycles. The terminology was introduced by Cayley, but the idea of a tree had been used ten years earlier by G. K. C. von Staudt [48] and by G. R. Kirchhoff [25], the latter using it in connection with his work on electrical networks. Moreover, the idea of a family tree goes back much earlier.

If one particular vertex is distinguished for some reason, then the tree is said to have that vertex as its *root*, and the structure itself is called a *rooted tree*. Cayley [2] considered the problem of finding the number A_n of rooted trees with n edges. If the root has r edges incident with it, then the removal from the tree of the root and these r edges produces r separate rooted trees. In the following figure, for example, the removal of the root ρ produces three separate rooted trees with roots ρ_1, ρ_2, and ρ_3.



Three rooted trees from one rooted tree.

Conversely, starting with a set of r rooted trees, we can form a single rooted tree by joining each of the old roots to a new vertex; the resulting structure may be regarded as a single rooted tree with the new vertex as its root.

From this simple idea, and mindful of the way in which Euler had constructed generating functions for partitions, Cayley was able to show that the generating function for rooted trees satisfies the following equation:

$$1 + A_1x + A_2x^2 + A_3x^3 + \cdots = (1-x)^{-1}(1-x^2)^{-A_1}(1-x^3)^{-A_2}(1-x^4)^{-A_3} \dots .$$

At a quick glance, this expression looks unhelpful for calculating the numbers A_n, but this is not so. If we expand the right-hand side, it turns out that the coefficient of x^n involves only the numbers $A_1, A_2, \dots, A_{n-1}$. So, for each n, the number A_n can be expressed in terms of its predecessors $A_1, A_2, \dots, A_{n-1}$ (with $A_0 = 1$), and the numbers can be calculated one at a time. Cayley did this as far as $n = 10$, but he made a few arithmetical errors, some of which he later corrected.

In 1859 Cayley published a second paper [3] in which he enumerated a special class of trees. Roughly speaking, the trees he considered were rooted trees with each *terminal vertex* (vertex of degree 1) at the same distance from the root. Denoting by p_n the number of such trees with n terminal vertices, he considered the 'exponential generating function'

$$P(x) = p_1 + \frac{p_2x}{1!} + \frac{p_3x^2}{2!} + \frac{p_4x^3}{3!} + \cdots ,$$

and showed that it satisfies the functional equation

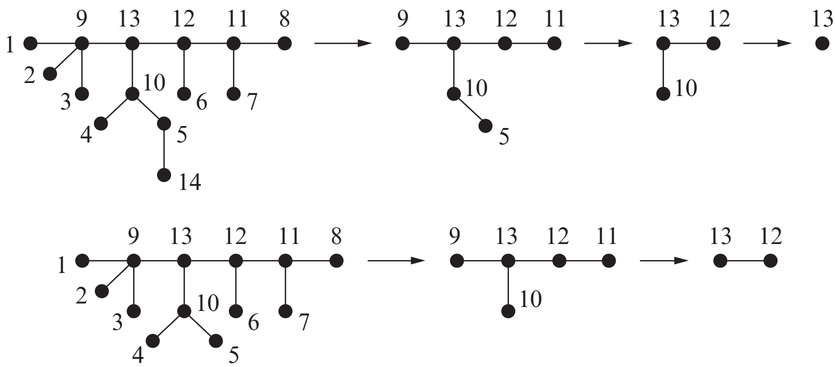
$$e^x P(x) = 2P(x) - 1.$$

Hence,

$$P(x) = \frac{1}{2 - e^x},$$

which can easily be expanded in powers of x to find p_n.

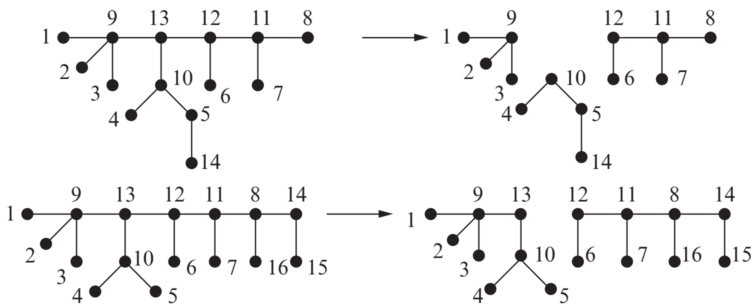
In 1869 Camille Jordan wrote a paper 'Sur les assemblages de lignes' [23], of which trees are a special case. If the outer edges of a tree are stripped off, a smaller tree is obtained, and if the process is repeated, one eventually obtains either a single vertex or a pair of vertices joined by an edge. In the former case, Jordan called the single vertex the *centre* of the original tree, and in the latter case he termed the two vertices of the edge the *bicentre*; the two cases are illustrated below. Using these concepts, Cayley was able to enumerate unrooted trees (see his brief note [4] and the more substantial [5]).



A tree with a centre, and one with a bicentre.

A different concept for the centrality of trees was also introduced by Jordan. Let v be any vertex of a tree T with n vertices, and let $T_1, T_2, \ldots, T_r$ be the subtrees obtained when v and its incident edges are removed. For most choices of v, one of the trees T_i contains more than $\frac{1}{2}n$ vertices, but if n is odd then there is a unique vertex, called the *centroid* of T, for which each of the trees has fewer than $\frac{1}{2}n$ vertices. If n is even, however, there may be either a unique vertex with this property, or a pair of vertices (necessarily joined by an edge) each of which has this property; in the latter case, the pair of vertices is called a *bicentroid*.

Removal of the edge joining the two vertices in a bicentroid produces two subtrees, each with $\frac{1}{2}n$ vertices. For example, in the first tree below, removing vertex 13 from the fourteen-vertex tree produces three subtrees, each with fewer than $\frac{14}{2} = 7$ vertices, but if any other vertex is removed, then one of the resulting subtrees has at least seven vertices; so vertex 13 is the centroid of the original tree. By contrast, if the edge joining vertices 13 and 12 in the second tree is removed, then each of the resulting subtrees has $\frac{16}{2} = 8$ vertices; so vertices 13 and 12 form the bicentroid.



A tree with a centroid, and one with a bicentroid.

Using these concepts Cayley was able to calculate the number of unrooted trees more easily (see [6]). His method was to regard an unrooted tree with a centroid as a rooted tree with the centroid as the root; for example, in the first tree above, vertex 13 is the root. Cayley regarded an unrooted tree with a bicentroid as composed of a pair of rooted trees with an edge joining the two roots; for example, if the two trees above are regarded as rooted trees with vertices 13 and 12 as the roots, then the first tree can be obtained by joining these vertices by an edge; this new edge and its end vertices form the bicentroid of the second tree.

In his calculations Cayley now worked in terms of the number ϕ_n of rooted trees with n vertices, rather than the number A_n with n edges, but since a tree with n vertices has $n - 1$ edges, $\phi_n = A_{n-1}$. The number β_n of bicentroidal trees (for n even) is easy to calculate, since to form such a tree with n vertices, one just selects a pair of rooted trees, each with $\frac{1}{2}n$ vertices, and joins the two roots. If the two choices are distinct, this can be done in $\frac{1}{2}\phi_{n/2}(\phi_{n/2} - 1)$ ways, but if the two choices are the same, this can be done in $\phi_{n/2}$ ways. The total number of unrooted bicentroidal trees with n vertices (n even) is thus

$$\tfrac{1}{2}\phi_{n/2}(\phi_{n/2} - 1) + \phi_{n/2} = \tfrac{1}{2}\phi_{n/2}(\phi_{n/2} + 1).$$

To form unrooted centroidal trees with n vertices each choice of a smaller rooted tree must have fewer than $\frac{1}{2}n$ vertices, and Cayley's method was to form a generating function in much the same way that Euler formed partition-generating functions. The number γ_n of unrooted centroidal trees is the coefficient of x_n in the expansion of

$$(1-x)^{-\gamma_1}(1-x^2)^{-\gamma_2}\cdots(1-x^{\lfloor n/2\rfloor})^{-\gamma_{\lfloor n/2\rfloor}}.$$

Finally, the number ϕ_n of rooted trees with n vertices is $\phi_n = \gamma_n$ (for n odd) and $\phi_n = \beta_n + \gamma_n$ (for n even).

Cayley also realized (see [4], [5]) that certain chemical enumeration problems could be reformulated in terms of enumerating specific classes of trees. For example, the carbon atoms and carbon–carbon bonds in an alkane C_nH_{2n+2} (also called a paraffin) form a tree with n vertices, in which each vertex has degree at most 4. The number of such trees is, therefore, equal to the number of isomers of C_nH_{2n+2}. The restriction on the degree is not easy to deal with, however, and although he published in this area, Cayley was not as successful here as he had been with the earlier tree-enumeration problems. Cayley's work on chemical enumeration was later followed up by George Pólya, whose work is discussed below. An account extending Cayley's work on alkanes is given by Rains and Sloane [37].

Major Percy MacMahon

Percy Alexander MacMahon (1854–1929) was born in Malta and had a military career, rising to the rank of major. He did see some active service in India, and later spent some years at the Royal Military Academy in Woolwich. Although he was part of the British mathematical establishment of the time (he became President of the London Mathematical Society in 1894–96 and of the Royal Astronomical Society in 1917–18), MacMahon's extensive researches in combinatorics rather set him apart from his contemporaries.

In his work MacMahon made great use of generating functions, and topics that he studied included partitions and symmetric functions (see Chapter 9). His collected papers were edited by Andrews [32], who added extensive commentaries on MacMahon's work.

MacMahon wrote two volumes entitled *Combinatory Analysis* [30], which are still in print, and these were soon followed by a short introduction [31] to the subject. In his preface to the latter, he explained that he was prompted to write it since some of his ‘mathematical critics’ had found that the original volumes were ‘difficult or troublesome reading’. Not everyone thought this, however, for Redfield was inspired by MacMahon’s books to do fundamental research into enumeration, as we now see.

J. Howard Redfield – a remarkable polymath

One of the most remarkable people to have worked in enumeration in the 20th century was the American polymath J. Howard Redfield. His educational qualifications included an SB degree from Haverford College, near Philadelphia, a second SB degree (in civil engineering) from the Massachusetts Institute of Technology, and a PhD degree (in Romance languages, with a thesis on ‘Romance loan-words in Basque’) from Harvard University. His professional work as a civil engineer led him to study the mathematical theory of elasticity from Augustus Love’s book [29], but he went on to study other mathematical material, including P. G. Tait’s work on knot theory (see [47, pp. 273–347]), Whitehead and Russell’s *Principia Mathematica* [50], and MacMahon’s *Combinatory Analysis* [30]. What prompted him to do so is not clear, but in following up some work of MacMahon, Redfield realized that enumerative combinatorics and group theory are interrelated, and he published a pioneering paper [42] on this borderline area in 1927. Unfortunately this paper was generally overlooked until the 1960s, by which time other researchers had independently obtained many of the results in Redfield’s paper.

Redfield’s work involved the idea of a group of symmetries. This is a collection of symmetries (such as rotations and reflections) that can be combined to give new symmetries. He realized that many enumeration problems can be specified in terms of placing n geometrical objects in n positions, where a group G of symmetries acts on the set of objects and a second symmetry group H acts on the set of positions. If G is a group of permutations of a set X of objects, then each element $g \in G$ can be expressed by combining cycles of objects, and Redfield realized that for enumeration problems it is only the numbers of cycles of the various lengths that are important, and not which elements are in which cycles.

Redfield introduced a polynomial Grf(G), associated with the group G acting on the set X, and called it the *group reduction function*; nowadays, it is usually termed the *cycle index* (see the next section). The polynomial may be written as

$$\mathrm{Grf}(G) = \frac{1}{|G|} \sum_{g \in G} s_1^{a_1(g)} s_2^{a_2(g)} s_3^{a_3(g)} \ldots,$$

where $a_1(g)$ is the number of objects fixed by the action of the symmetry g, $a_2(g)$ is the number of 2-cycles (pairs of objects interchanged by g), $a_3(g)$ is the number of 3-cycles, and so on. Here, the symbols s_r can be regarded as variables or indeterminates, although Redfield regarded them as the power-sum symmetric functions

$$s_r = x_1^r + x_2^r + \cdots + x_n^r$$

in the indeterminates $x_1, x_2, \ldots, x_n$.

Redfield also introduced a composition (binary operation) on polynomials which, in the type of problem mentioned above, he applied to $\mathrm{Grf}(G)$ and $\mathrm{Grf}(H)$. Here (following Read [39], [40]), the composition is denoted by $*$. The rules defining $*$ are:

- for identical monomials,

$$s_1^a s_2^b s_3^c \cdots * s_1^a s_2^b s_3^c \cdots = 1^a a! 2^b b! 3^c c! \ldots s_1^a s_2^b s_3^c \ldots;$$

- the composition of two non-identical monomials is 0;
- the composition of polynomials is linear – equivalently, $*$ is distributive over $+$.

One example, considered by Redfield, is to find the number of ways of placing four black balls and four white balls at the corners of a cube, with one ball at each corner. Here, the four black balls can be permuted in any way, and so can the four white ones. Thus, the object group G is obtained by combining two copies of the group S_4 of all permutations of four objects, and its group reduction function is the square of that for the symmetric group S_4, so

$$\begin{aligned}\mathrm{Grf}(G) &= \frac{1}{24^2}\left(s_1^4 + 8s_1s_3 + 6s_1^2s_2 + 3s_2^2 + 6s_4\right)^2 \\ &= \frac{1}{24^2}\left(s_1^8 + 8^2s_1^2s_3^2 + 3^2s_2^4 + 6^2s_4^2 + \cdots\right).\end{aligned}$$

The position group H is the group of all rotational symmetries of the cube, with

$$\mathrm{Grf}(H) = \tfrac{1}{24}\left(s_1^8 + 8s_1^2s_3^2 + 9s_2^4 + 6s_4^2\right).$$

The solution to the enumeration problem is obtained by composing these two group reduction functions and then adding the coefficients. Because of the second rule above, the only monomials that contribute to the composition are those that occur in each group reduction function, so

$$\mathrm{Grf}(G) * \mathrm{Grf}(H) = \frac{1}{24^3}\Big((s_1^8 * s_1^8) + (8^2 s_1^2 s_3^2 * 8 s_1^2 s_3^2) + (3^2 s_2^4 * 9 s_2^4) + (6^2 s_4^2 * 6 s_4^2)\Big).$$

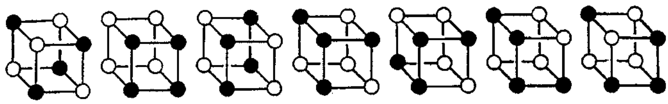
The sum of the coefficients in this composition is

$$\frac{1}{24^3}\Big(1^8 8! + (512 \times 1^2 2! 3^2 2!) + (81 \times 2^4 4!) + (216 \times 4^2 2!)\Big) = 7.$$

Thus, there are seven ways of placing the balls in the corners (see below).

REDFIELD: *The Theory of Group-Reduced Distributions.* 443

The actual configurations, shown below, cannot be determined by the methods of the present theory, but must be found, as in all other cases, by detailed consideration of the groups involved, and this may of course be very laborious, except in simple cases, or where special devices are available.



(1) (2) (3) (4) (5) (6) (7)

In connection with the present example we may note without proof certain other simple results obtainable.

Thus if in V we substitute $x^r + y^r$ for every s_r, we obtain the polynomial

$$x^8 + x^7y + 3x^6y^2 + 3x^5y^3 + 7x^4y^4 + 3x^3y^5 + 3x^2y^6 + xy^7 + y^8,$$

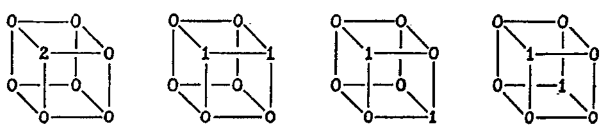
in which the coefficient of $x^t y^{8-t}$ enumerates the distinct configurations possible with t nodes • and $8 - t$ nodes °.

The sum of the coefficients in the above expression is 23, which is the total number of configurations when the numbers of nodes of the two colors are not specified. This enumeration is also effected by substituting 2 for every s_r in V. Similarly if k colors are available we substitute k for every s_r; thus with 3 colors there are $(1/24)(3^8 + 9.3^4 + 8.3^4 + 6.3^2) = 333$ possible configurations.

If in V we put $1/(1 - x^r)$ for every s_r, we obtain the infinite series

$$1 + x + 4x^2 + 7x^3 + 21x^4 + 37x^5 + \cdots,$$

in which the coefficient of x^t enumerates the distinct configurations obtained by placing a zero or a positive integer at every vertex of the cube, subject to the condition that the sum of the 8 numbers is always t. For $t = 2$, the 4 configurations are



If in V we put 2 for every s_{2k} and 0 for every s_{2k+1}, we enumerate the configurations in which it is possible to change the color of every node into

A page from Redfield's paper, solving the above problem.

The above problem can be restated in terms of establishing a one–one correspondence between balls and corners. When this point of view is adopted, the two sets (objects and positions) play more symmetrical roles, and it is an easy step to generalize to one–one– $\cdots$ –one correspondences between q sets of objects, where each set has a group G_i of symmetries (for $i = 1, 2, \ldots, q$). This Redfield did, and his calculations involved the composition $\mathrm{Grf}(G_1) * \mathrm{Grf}(G_2) * \cdots * \mathrm{Grf}(G_q)$ of all the group reduction functions.

Redfield also noted that each distribution has its own symmetry group (which is necessarily contained in the position group H), and he asked whether it is possible to break down the counting so as find the number n_i of distributions with a given symmetry group $H_i \subseteq H$. He showed that the composition $\mathrm{Grf}(G) * \mathrm{Grf}(H)$ satisfies

$$\mathrm{Grf}(G) * \mathrm{Grf}(H) = \sum_i n_i \ \mathrm{Grf}(H_i).$$

Unfortunately, different G_i may sometimes have the same Grf, or the various Grfs may be linearly dependent, and in such cases the numbers n_i cannot be determined uniquely from the above equation. Thus, in general, the group reduction function is not the appropriate tool to find the numbers n_i. Redfield continued his researches, however, and by 1937 he had found the right tool.

Redfield carried out much of his mathematical research at his home on Farm Road in Wayne, Pennsylvania, about fifteen miles from central Philadelphia. He had some contact with professional mathematicians in the area, and also access to libraries in Philadelphia, but often material that would have helped him was not readily available to him. In particular, he realized that group characters were relevant to his work, but he was unable to see much of the literature on that subject. The tool that Redfield eventually discovered to find the numbers n_i he called a *generalized character*, but it was already in the literature in the second edition of Burnside's book [1] under the name *mark* (of a permutation group). It seems, however, that Redfield never saw the second edition of that book, but only the first edition.

In 1937 Redfield gave a lecture on his work at the University of Pennsylvania. The typescript that he prepared for the lecture has survived and was published in 2000 (see Redfield [43] and Lloyd [28]). It is a good introduction to Redfield's ideas and methods, and includes his work on generalized characters (marks). Finally, in 1940, Redfield submitted a second paper to the *American Journal of*

Mathematics, but alas it was rejected. It was not to appear in print until 1984 [44], some forty years after his death.

The Redfield family has preserved many of J. Howard Redfield's notebooks and manuscripts, including a copy of a letter to D. E. Littlewood, dated 9 May 1938. In this letter, Redfield thanked Littlewood for sending him some offprints, and he went on to make some remarks about their contents. It is not known who initiated the correspondence, nor whether there were any further letters, but it helps to explain the fact that Littlewood briefly cited Redfield's first paper [42] in both editions of his book [26]. This appears to be the only citation of Redfield's paper during his lifetime, and no other citation is known before a glowing account of Redfield's achievements by Frank Harary [18] in 1960. Since then, several analyses and numerous citations of Redfield's work have appeared (see, in particular, Hall, Palmer, and Robinson [17] and Sheehan [46]). For further details of Redfield's life and the discovery of his unpublished researches, see Lloyd [27], [28].


J. Howard Redfield (1879–1944) and George Pólya (1887–1985).

The work of Pólya

In the mid 1930s George Pólya, who had already established a reputation in analysis, wrote a few short papers on enumeration. He then wrote a very lengthy paper [35] which was quickly recognized as a landmark in the subject, but it was to be many years before an English translation of it was published, in [36]. Extending Cayley's work, Pólya enumerated various types of graphs and chemical compounds. The concept that enabled him to progress further than Cayley was a polynomial which he termed the *cycle index*; it was just Redfield's group reduction function under another name, but was undoubtedly conceived independently by Pólya.

Unlike Redfield, Pólya did not combine different cycle indices. Instead, each of his calculations involved only a single cycle index in which the variables are replaced by generating functions (usually infinite series) to produce a new generating function. (It is clear from the extract on p. 300 from Redfield's 1927 paper that he too was aware that cycle indices could be used in this way.) In spite of its length, Pólya's paper is largely based on a single theorem, his 'Hauptsatz', which he applied to a great many examples.

To illustrate Pólya's technique, we return to our earlier problem of placing black and white balls at the corners of a cube. This can be solved in the following alternative way. The polynomial $\phi(b, w) = b + w$ is regarded as a generating function for the black and white balls and, as mentioned above, the cycle index of the cube acting on its vertices is

$$\mathrm{Grf}(H) = \tfrac{1}{24}\left(s_1^8 + 8s_1^2 s_3^2 + 9s_2^4 + 6s_4^2\right).$$

The substitutions $s_n = b^n + w^n$ are then made into the cycle index to produce the generating function

$$\begin{aligned}\Phi(b, w) &= \tfrac{1}{24}\left((b+w)^8 + 8(b+w)^2(b^3+w^3)^2 + 9(b^2+w^2)^4 + 6(b^4+w^4)^2\right)\\ &= b^8 + b^7 w + 3b^6 w^2 + 3b^5 w^3 + 7b^4 w^4 + 3b^3 w^5 + 3b^2 w^6 + bw^7 + w^8.\end{aligned}$$

Here, the coefficient 7 of $b^4 w^4$ is the number of ways of placing four black balls and four white balls at the corners of the cube. More generally, the coefficient of $b^r w^s$ is the number of ways of placing r black balls and s white balls at the corners of the cube; this is necessarily 0 unless $r + s = 8$. Thus Pólya's method of solving the original problem simultaneously solves the analogous problem for all possible distributions of balls.

Over the years, many people have used Pólya's theorem to count various families of objects. It is impossible to give an exhaustive list here, but the book by Harary and Palmer [20] includes many examples of its use in graph theory.

Later work

In the 1950s, in his PhD thesis [38] (see also [39], [40]), Ronald C. Read introduced his 'superposition theorem', which involved the composition of cycle indices; he was, without realizing it, rediscovering the technique used by Redfield some thirty years earlier. Other developments were made in both the theory and the applications of enumeration during the 1960s and 1970s. Researchers involved included Herbert Foulkes, N. G. de Bruijn, John Sheehan, and Frank Harary and his co-workers.

De Bruijn reformulated Pólya theory in terms of counting equivalence classes of functions $f : D \to R$, where D and R are sets. If a symmetry group G acts on D, two functions f_1 and f_2 are *equivalent* whenever there is an element $g \in G$ for which $f_1 g = f_2$ – that is, for all $d \in D, f_1 g(d) = f_2(d)$. For the earlier problem of placing black and white balls at the corners of a cube, the set D is the set of corners, $R = \{b, w\}$ is the set of two colours, and G is the rotation group of the cube acting on its corners.

De Bruijn also generalized the theory in various ways; for example, he introduced a second group H acting on R, and defined two functions f_1 and f_2 to be equivalent if there are an element $g \in G$ and an element $h \in H$ for which $f_1 g = hf_2$. For the problem above with two colours, the only non-trivial possibility for H is the group of order 2, where the non-identity element interchanges the two colours b and w; this corresponds to counting patterns of colours, where the actual colours used are of no consequence. Thus, a colouring using r white balls and s black balls would not be counted separately from the corresponding one where all the white balls had been replaced by black ones and all the black balls had been replaced by white ones.

De Bruijn's theorems involve the application of differential operators to cycle indices, but the calculations are often difficult and comparatively little use has been made of his methods. De Bruijn's paper [7] is a good introduction to the simpler parts of his work; for later generalizations, see [8].

There is an intimate connection between the theorems of Pólya, Redfield, and de Bruijn, and in the case of the earlier de Bruijn generalizations, for example, it is possible to replace the two groups by a single composite group acting on a different set. Pólya's theorem can then be applied to solve the original problem, and this obviates the need to use differential operators (see Harary and Palmer [19], [20], for details).

The connection between Redfield's 1927 paper and group characters was developed by Herbert Foulkes [13], [14]. His student, John Sheehan, attacked the problem (mentioned earlier) of counting distributions according to their individual symmetry groups. In [45] he solved it using marks of permutation groups, unaware that Redfield had already done this in the late 1930s, since Redfield's solution [44] was not published until 1984.

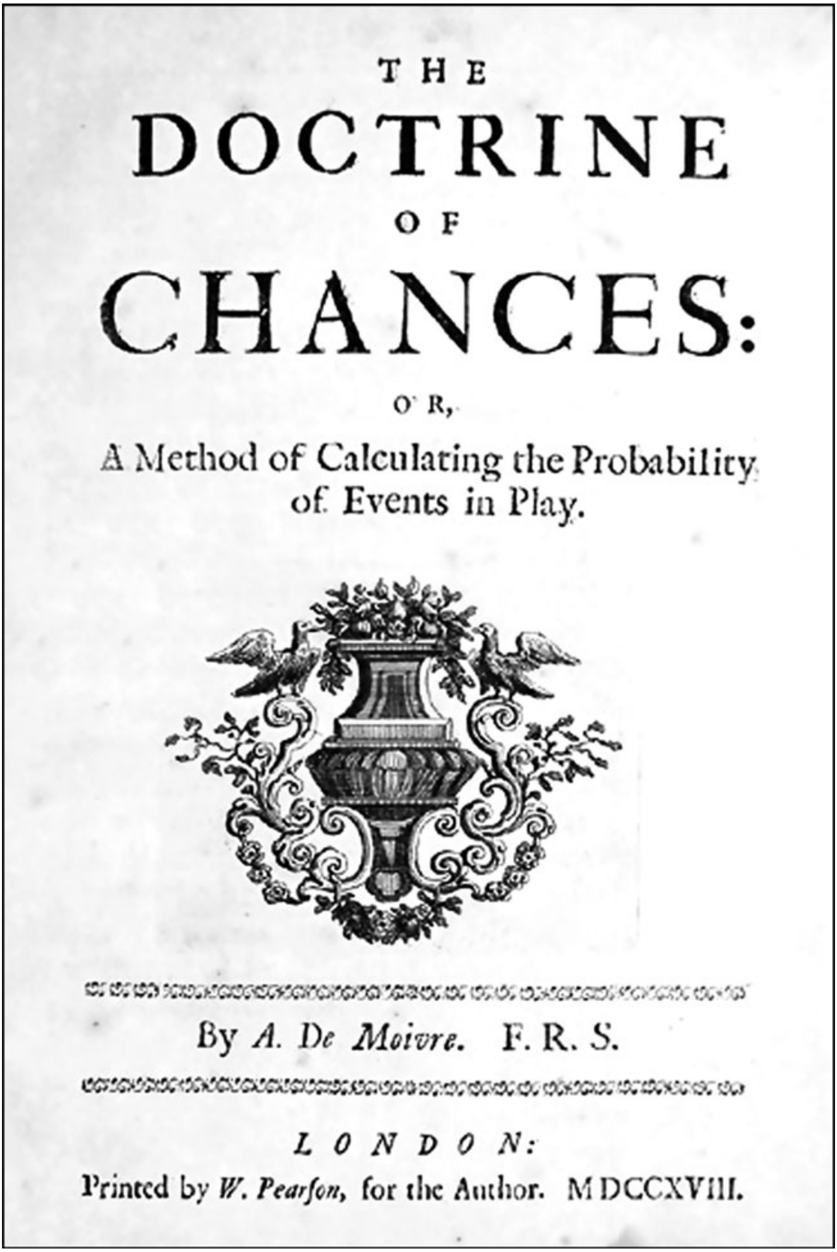
For information on more recent work in the field of enumeration, see Read [41] and Kerber [24] for accounts by mathematicians, and Fujita [15] and El-Basil [11] for accounts by chemists.

References

1. W. Burnside, *Theory of Groups of Finite Order*, Cambridge University Press (1897); 2nd edn. (1911) (reprinted by Dover, 1955).
2. A. Cayley, On the theory of the analytical forms called trees, *Philos. Mag.* (4) 13 (1857), 172–6 = *Math. Papers* 3, 242–6.
3. A. Cayley, On the theory of the analytical forms called trees. Second part, *Philos. Mag.* (4) 18 (1859), 374–8 = *Math. Papers* 4, 112–15.
4. A. Cayley, Über die analytischen Figuren, welche in der Mathematik Bäume genannt werden und ihre Anwendung auf die Theorie chemischer Verbindungen, *Ber. Deutsch. Chem. Ges.* 8 (1875), 1056–9.
5. A. Cayley, On the analytical forms called trees, with applications to the theory of chemical combinations, *Rep. Brit. Assoc. Adv. Sci.* 45 (1875), 257–305 = *Math. Papers* 9, 427–60.
6. A. Cayley, On the analytical forms called trees, *Amer. J. Math.* 4 (1881), 266–8 = *Math. Papers* 11, 365–7.
7. N. G. de Bruijn, Pólya's theory of counting, *Applied Combinatorial Mathematics* (ed. E. F. Beckenbach), Wiley (1964), 144–184.
8. N. G. de Bruijn, A survey of generalizations of Pólya's enumeration theorem, *Nieuw Archief Wisk.* (2) 19 (1971), 89–112.
9. A. de Moivre, A method of raising an infinite multinomial to any given power, or extracting any given root of the same, *Philos. Trans.* 230 (1697), 619–25.
10. A. de Moivre, *Miscellanea Analytica de Seriebus*, Book iv, London (1730).
11. S. El-Basil, *Combinatorial Organic Chemistry: An Educational Approach*, Nova Science (2000).
12. W. Emerson, *The Method of Increments*, J. Nourse (1763).
13. H. O. Foulkes, On Redfield's group reduction functions, *Canad. J. Math.* 15 (1963), 272–84.
14. H. O. Foulkes, On Redfield's range correspondences, *Canad. J. Math.* 18 (1966), 1060–71.

15. S. Fujita, *Symmetry and Combinatorial Enumeration in Chemistry*, Springer (1991).
16. K. Haas, Hindenburg, *Dictionary of Scientific Biography*, Vol. VI, Charles Scribner's Sons (1972), 403–4.
17. J. I. Hall, E. M. Palmer, and R. W. Robinson, Redfield's lost paper in a modern context, *J. Graph Theory* 8 (1984), 225–40.
18. F. Harary, Unsolved problems in the enumeration of graphs, *Publ. Math. Inst. Hungarian Acad. Sci.* 5 (1960), 63–95; Addendum 1 to this paper discusses Redfield's work.
19. F. Harary and E. M. Palmer, The power group enumeration theorem, *J. Combin. Theory* 1 (1966), 157–73.
20. F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press (1973).
21. C. F. Hindenburg, *Der polynomische Lehrsatz, das wichtigste Theorem der ganzen Analysis* . . . , G. Fleischer, Jr. (1796).
22. C. F. Hindenburg, *Sammlung Combinatorisch–Analytischer Abhandlungen*, G. Fleischer, Jr. (1800).
23. C. Jordan, Sur les assemblages de lignes, *J. Reine Angew. Math.* 70 (1869), 185–90 = *Oeuvres* 4, 303–8.
24. A. Kerber, *Applied Finite Group Actions*, 2nd edn., Springer (1999).
25. G. R. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilungen galvanischer Ströme geführt wird, *Ann. Phys. Chem.* 72 (1847), 497–508 = *Ges. Abh.*, 22–33.
26. D. E. Littlewood, *The Theory of Group Characters*, Oxford University Press (1940); 2nd edn. (1950).
27. E. K. Lloyd, J. Howard Redfield 1879–1944, *J. Graph Theory* 8 (1984), 195–203.
28. E. K. Lloyd, Redfield's 1937 lecture, *MATCH, Commun. Math. Comput. Chem.* 41 (2000), 29–41.
29. A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, 2 vols., Cambridge University Press (1892–3).
30. P. A. MacMahon, *Combinatory Analysis*, 2 vols., Cambridge University Press (1915–16) (reprinted by Chelsea, 1960).
31. P. A. MacMahon, *Introduction to Combinatory Analysis*, Cambridge University Press (1920) (reprinted in *Famous Problems and Other Monographs*, Chelsea, 1955; 2nd edn. 1962).
32. P. A. MacMahon, *Percy Alexander MacMahon: Collected Papers* (ed. G. E. Andrews), 2 vols., MIT Press (1978, 1986).
33. P. Nicholson, *An Introduction to the Method of Increments*, Scientific Press (1817).
34. P. Nicholson, *Essays on the Combinatorial Analysis*, Longman, Hurst, Rees, Orme, and Brown (1818).
35. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* 68 (1937), 145–254.
36. G. Pólya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*, Springer (1987) (includes an English translation of [35]).
37. E. M. Rains and N. J. A. Sloane, On Cayley's enumeration of alkanes (or 4-valent trees), *J. Integer Sequences* 2 (1999), Article 99.1.1.

38. R. C. Read, *Some Enumeration Problems in Graph Theory*, PhD thesis, University of London (1958).
39. R. C. Read, The enumeration of locally restricted graphs, *J. London Math. Soc.* 34 (1959), 417–36.
40. R. C. Read, The enumeration of locally restricted graphs II, *J. London Math. Soc.* 35 (1960), 344–51.
41. R. C. Read, The legacy of Pólya's paper: 50 years of Pólya theory, in [36], 96–135.
42. J. H. Redfield, The theory of group-reduced distributions, *Amer. J. Math.* 49 (1927), 433–55 (reprinted in [32], Vol. 1, 805–27).
43. J. H. Redfield, Group theory applied to combinatory analysis (typescript of a lecture delivered at the University of Pennsylvania, 1937), *MATCH Commun. Math. Comput. Chem.* 41 (2000), 7–27.
44. J. H. Redfield, Enumeration by frame group and range group (written in 1940), *J. Graph Theory* 8 (1984), 205–23.
45. J. Sheehan, The number of graphs with a given automorphism group, *Canad. J. Math.* 20 (1968), 1068–76.
46. J. Sheehan, Redfield discovered again, *Surveys in Combinatorics* (ed. E. K. Lloyd), London Mathematical Society Lecture Note Series 82, Cambridge University Press (1983).
47. P. G. Tait, *Scientific Papers* 1, Cambridge University Press (1898).
48. G. K. C. von Staudt, *Geometrie der Lage*, Fr. Korn'schen Buchhandlung (1847).
49. I. C. Weingärtner, *Lehrbuch der kombinatorischen Analysis, nach der Theorie des Herrn Professor Hindenburg ausgearbeitet*, 2 vols, Leipzig (1800–1).
50. A. N. Whitehead and B. Russell, *Principia Mathematica*, 3 vols., Cambridge University Press (1910–13).



THE

DOCTRINE

OF

CHANCES:

OR,

A Method of Calculating the Probability of Events in Play.

By *A. De Moivre.* F. R. S.

LONDON:

Printed by *W. Pearson,* for the Author. MDCCXVIII.

Title page of Abraham de Moivre's *The Doctrine of Chances.*

CHAPTER 13

Combinatorial set theory

IAN ANDERSON

This chapter outlines the historical development of the study of combinatorial problems concerning finite sets, beginning with the inclusion–exclusion principle of de Moivre in the early 18th century, and finishing with the 20th-century development of a unified body of theory relating to the intersections, unions, and orderings of collections of finite sets.

Introduction

A *set* is a collection of objects. The generality of this concept might suggest that little of interest can be said about sets, but this is far from being the case. Indeed, mathematics is largely concerned with sets – the set of numbers, the set of lines and points in the plane, the set of matrices, and so on.

Often we are concerned with the size of a set:

How many prime numbers are there in the set $\{1, 2, \ldots, 1000\}$?

How many elements of a set have a given property?

How large can a collection of subsets of a set be if no subset is contained in another?

In this chapter we explore several key ideas that have arisen during the past three hundred years and which are of interest to mathematicians today. These are mainly concerned with set-theoretic concepts such as intersections and set inclusion, although the contexts vary greatly.

The first topics in this chapter are two basic counting principles called the *inclusion–exclusion principle* and the *box principle*. The first of these is concerned with estimating the size of the union of given subsets of a set, and we present a general formula that gives the answer in terms of the sizes of the intersections of the subsets. This principle was first exploited by Abraham de Moivre in 1718 in his book *The Doctrine of Chances* [4] (see Chapter 6), although he did not use abstract set notation in his presentation. He was concerned with the number of permutations of objects in which certain places are forbidden to certain of the objects; this *derangement problem* was solved by a general method that was later abstracted from his concrete problems as the *inclusion–exclusion principle* and applied more generally.

The second principle, the *box principle* or *pigeonhole principle*, essentially asserts that if more than *n* objects are placed into *n* boxes, then at least one box must receive at least two objects. This simple idea was developed almost beyond recognition in the 20th century, by the logician F. P. Ramsey and others, to prove the existence of substructures of a required type in a larger structure, provided that the original structure is sufficiently large.

The problem of selecting different elements to 'represent' given subsets is fundamental to many problems in set theory and graph theory. The origin of this problem of finding *distinct representatives* was in the area of finite geometry at the end of the 19th century, but the same problem reappeared in the work of 20th-century mathematicians such as Dénes König and Philip Hall. The basic problem is popularly known as the *marriage problem*, and we describe its development later in the section on tranversals. The story is one of independent discoveries, in different contexts, of results that turned out to be equivalent, although formulated in very different ways.

The second half of the 20th century saw a great increase of interest in the study of families of subsets of a set that possess certain intersection or inclusion properties. *Combinatorial set theory*, as it has come to be called, is often considered to have begun with a theorem of Sperner in 1928, which put a bound on the number of subsets of a set that can be chosen if no one subset is to be contained in any other. How subsets relate to one another under inclusion parallels how numbers are related under division, and these problems can be viewed as special cases of problems arising in general partially ordered sets. We study *Sperner's theorem* and show how it was used by later authors to solve other set-theoretic problems. Finally, we discuss the later results of Erdős, Ko, and Rado, and of Kruskal and Katona, which have proved central to the development of the subject.

The inclusion–exclusion principle

In the 24th and 25th problems, I explain a new sort of algebra, whereby some questions relating to combinations are solved by so easy a process that their solution is made in some measure an immediate consequence of the method of notation. I will not pretend to say that this new algebra is absolutely necessary for the solving of these questions which I make to depend on it, since it appears that Mr Montmort, author of *Analyse des Jeux de Hazard*, and Mr Nicholas Bernoulli have solved, by another method, many of the cases therein proposed: but I hope I shall not be thought guilty of too much confidence if I assure the reader that the method I have followed has a degree of simplicity, not to say of generality, which will hardly be attained by any other steps than by those I have taken.

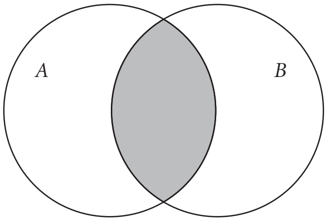
So wrote Abraham de Moivre in the preface to his book *The Doctrine of Chances* [4], published in 1718. De Moivre, a Frenchman by birth, came to England in 1685 as a Protestant refugee, spending the rest of his life in London; he became a Fellow of the Royal Society in 1697. Clearly de Moivre had read the 1708 book *Essay d'Analyse sur les Jeux de Hazard* (Essay on the Analysis of the Games of Chance) [5] by Pierre Rémond (later de Montmort). This book was a study of games of chance, and included Pascal's triangle and the derangement problem, which de Moivre was to solve by his own new method. The derangement problem asks:

If a number of objects in order are arbitrarily rearranged, how likely is it that no object will be in its original position?

De Moivre's 'new sort of algebra' is essentially a generalization of the answer to the following simple set-theoretic question:

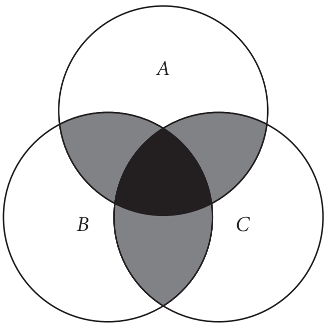
Given two finite sets A and B, how big is their union $A \cup B$?

If $|X|$ denotes the number of elements in the set X, we observe that the estimate $|A| + |B|$ may be too big, since any elements in both sets are counted twice, and so we refine our estimate by subtracting the number of elements in both sets (represented by the shaded area in the diagram). First we *include* elements in A or in B, and then we *exclude* those in both. Thus:



$$|A \cup B| = |A| + |B| - |A \cap B|.$$

This idea extends to three sets, A, B, and C:



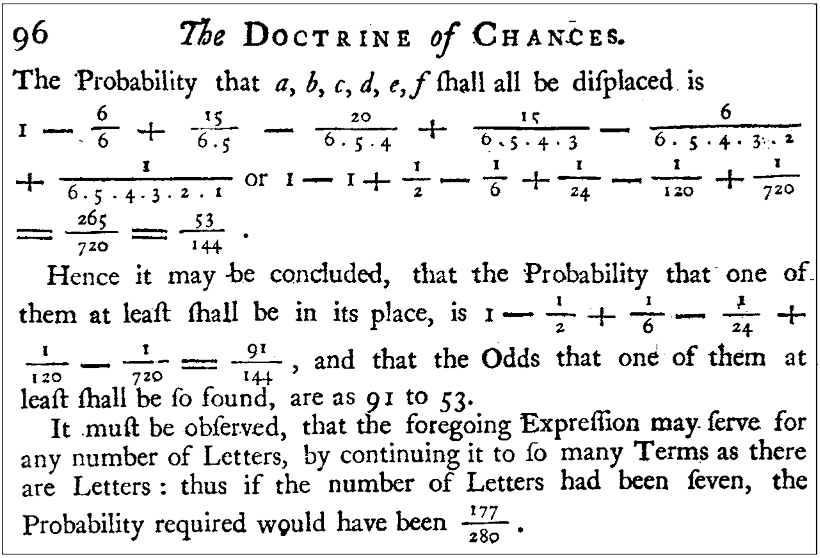
$$\begin{aligned}|A \cup B \cup C| = {} & |A| + |B| + |C| \\ & - |A \cap B| - |A \cap C| - |B \cap C| \\ & + |A \cap B \cap C|.\end{aligned}$$

Here we include and exclude three times any element that is in all three sets, so we have to include such elements once again at the end.

De Moivre argued that if A is the set of arrangements of a, b, c, d, e, f in which a is in its correct place, and if B, C, D, E, F are similarly defined, then the *number of derangements* (orderings in which no letter is in its correct place) is

$$\begin{aligned}& 6! - |A \cup B \cup \ldots \cup F| \\ & = 6! - \Big\{ \big(|A| + |B| + \cdots + |F|\big) - \big(|A \cap B| + \cdots + |E \cap F|\big) + \big(|A \cap B \cap C| + \cdots\big) \\ & - \big(|A \cap B \cap C \cap D| + \cdots\big) + \big(|A \cap B \cap C \cap D \cap E| + \cdots\big) - |A \cap B \cap C \cap D \cap E \cap F| \Big\} \\ & = 6! - C(6,1)5! + C(6,2)4! - C(6,3)3! + C(6,4)2! - C(6,5)1! + C(6,6) = 265.\end{aligned}$$

The *probability* that no object will be in its original position is then obtained by dividing by $6! = 720$, giving $\frac{265}{720} \approx 0.3680$, which is very close to $1/e \approx 0.3679$; in fact, the number of derangements of a set of n elements is always the integer nearest to $n!/e$, which in this case is $6!/e \approx 264.873$.



96 *The* DOCTRINE *of* CHANCES.

The Probability that *a*, *b*, *c*, *d*, *e*, *f* ſhall all be diſplaced is $1 - \frac{6}{6} + \frac{15}{6.5} - \frac{20}{6.5.4} + \frac{15}{6.5.4.3} - \frac{6}{6.5.4.3.2} + \frac{1}{6.5.4.3.2.1}$ or $1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} + \frac{1}{720} = \frac{265}{720} = \frac{53}{144}$.

Hence it may be concluded, that the Probability that one of them at leaſt ſhall be in its place, is $1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24} + \frac{1}{120} - \frac{1}{720} = \frac{91}{144}$, and that the Odds that one of them at leaſt ſhall be ſo found, are as 91 to 53.

It muſt be obſerved, that the foregoing Expreſſion may ſerve for any number of Letters, by continuing it to ſo many Terms as there are Letters: thus if the number of Letters had been ſeven, the Probability required would have been $\frac{177}{280}$.

De Moivre's use of the inclusion–exclusion principle.

The principle used by de Moivre is now called the *inclusion–exclusion principle*. In the past it has been known by various names, such as the *cross classification principle*, and has been used by many mathematicians over the centuries. It was, for example, used by Whitney [44] in 1932 in his study of the chromatic number of a graph, and it can also be used to obtain the 1998 formula of Ollerenshaw and Brée [35] for the number of most-perfect pandiagonal magic squares.

The inclusion–exclusion principle has also proved itself to be a very useful technique in the study of prime numbers. Eratosthenes of Cyrene (276–194 BC) may be best known for his remarkable measurement of the Earth's circumference, using the length of the shadow of a pole at noon in two different locations, but he is also remembered for his 'sieve' which he used to count prime numbers. To find all the prime numbers up to 100, for example, score out all numbers up to 100 that are multiples of 2, then score out all multiples of 3, 5, and 7; the remaining numbers, apart from 1, are the required primes, since every non-prime number less than 100 must have a prime factor less than 10. If this procedure is used to count primes, there is the added complication that some numbers are scored out more than once, so the inclusion–exclusion principle is appropriate. For numbers much larger than 100 there are problems with this method, arising from the large number of terms involved; these problems were overcome in 1920 by the Norwegian mathematician Viggo Brun, whose deep and clever adaptation [2] of the principle, known as the *Brun sieve*, has led to much progress in the study of the distribution of prime numbers.

The pigeonhole principle and Ramsey theory

In 1801, at the age of 24, Carl Friedrich Gauss published his great work *Disquisitiones Arithmeticae* [17]. In Article 45 of this book, Gauss considered the geometric progression $1, a, a^2, \ldots$ (modulo p):

Since the modulus p is prime relative to a and hence to any power of a, no term of the progression will be $\equiv 0 \pmod p$ *but each of them will be congruent to one of the numbers* $1, 2, \ldots, p-1$. *Since the number of these is* $p-1$, *it is clear that if we consider more than* $p-1$ *terms of the progression, not all can have different least residue. So among the terms* $1, a, a^2, \ldots, a^{p-1}$, *there must be at least one congruent pair ….*

Gauss is giving here one of the first known applications of the *pigeonhole principle*:

If $m+1$ *objects are placed in m pigeonholes, then at least one of the pigeonholes receives at least two objects.*

Thus, for example, in any group of thirteen people, there must be at least two people with birthdays in the same month. This simple idea has many applications. It is often called *Dirichlet's box principle*, since it was used by Dirichlet in his work [8] on the approximation of irrational numbers by rationals. It can also be used to show simply that a real number has a finite or recurring decimal expansion if and only if it is rational. A more general form of the principle is:

If $mn+1$ *objects are placed in m pigeonholes, then at least one of the pigeonholes receives at least* $n+1$ *objects.*

This section deals with some of the profound generalizations of this simple principle. An example appears below.

In 1916 the algebraist Issai Schur wrote a paper [37] on the congruence

$$x^m + y^m \equiv z^m \pmod p.$$

In that paper he gave a simple proof of a result of L. E. Dickson [6], that this congruence must be satisfied by three integers x, y, z that are relatively prime

Dirichlet's 'boxes' in the 17th century

Pierre Nicole, one of the authors of the famous *Logique de Port-Royal* (1662), related the following: 'One day I told Madame de Longueville that I could prove that there are at least two people living in Paris with the same number of hairs on their heads. She asserted that I could never prove this without counting them first. My premisses are these. No head has more than 200 000 hairs, and the worst case provided has one. Consider 200 000 heads, none having the same number of hairs. Then each must have a number of hairs equal to some number from 1 to 200 000 both included. Of course if any have the same number of hairs my bet is won. Now take one more person, who has not more than 200 000 hairs on his head. His number must be one of the numbers 1 to 200 000 included. As the inhabitants of Paris are nearer 800 000 than 200 000, there are many heads with an equal number of hairs.'

to p, provided that p is large enough. This result of Dickson showed that any attempt to prove Fermat's last theorem by considering the equation as a congruence modulo a prime is doomed to failure. Schur used the pigeonhole principle to show that, if n is sufficiently large and if the integers $1, 2, \ldots, n$ are distributed into k classes, then there must be three integers x, y, z, all in the same class, for which $x + y = z$. From this, Schur was able to give his simple proof of Dickson's result.

A related result was published in 1927 by B. L. van der Waerden [41]. Instead of considering the equation $z - y = x$, consider the system of equations

$$x_1 - x_2 = x_2 - x_3 = \cdots = x_{h-1} - x_h.$$

Here we require that the x_i are in arithmetic progression. Van der Waerden proved that, if the integers $1, 2, \ldots, n$ are distributed into k classes where n is sufficiently large, then there exist integers $x_1, x_2, \ldots, x_h$ in arithmetic progression and all in the same class.

A profound development of the pigeonhole principle appeared in 1930 in a posthumous paper by the brilliant young logician F. P. Ramsey. Ramsey, the elder brother of the future Archbishop of Canterbury, was born in Cambridge and was educated at Winchester and at Trinity College, Cambridge. Elected to a Fellowship at King's College in 1924, he taught there until his early death in

1930, caused by a liver infection after an abdominal operation, in a London hospital.

On 13 December 1928 Ramsey read a paper at a meeting of the London Mathematical Society on the *Entscheidungsproblem* (decision problem) of Hilbert, which was concerned with the possibility of finding a systematic method for determining the truth or falsity of a given logical formula. Ramsey's partial solution of this problem (which was completely solved by Alonzo Church in 1936) required a deep extension of the pigeonhole principle. Instead of distributing *elements* of an n-element set S into boxes, Ramsey distributed the *subsets* of a given size r into boxes; we can think of 'colouring' the subsets, each subset in a particular box being given the same colour. Ramsey showed [36] that, provided n is large enough, if each r-element subset of S is given one of k different colours then, for a given $l < n$, there must be a subset W of S of size l *all* of whose r-element subsets have the same colour.

In a sense Ramsey's theorem asserts that, in a random chaotic structure, there must be a certain amount of non-chaotic substructure somewhere within it. A well-known simple case of Ramsey's result occurs in the following problem, which appeared (see [18]) in the 1953 William Lowell Putnam Mathematical Competition (an annual competition sponsored by the Mathematical Association of America):

Six points are in general position in space (no three in line, no four in a plane). The fifteen line segments joining them in pairs are drawn and then painted, some segments red, some blue. Prove that some triangle has all its sides the same color.

This is the same problem as the following version that appeared as elementary problem E1321 on page 446 of the *American Mathematical Monthly* in 1958:

Prove that, at a gathering of any six people, some three of them are either mutual acquaintances or complete strangers to one another.

The problem is easily solved. Take any one of the points and consider the five lines from it. By the pigeonhole principle, at least three of these five lines must have the same colour. Without loss of generality, suppose the points are labelled $a, b, \ldots, f$ and that the lines ab, ac, and ad are all blue. Then, if any of the lines bc, cd, db is blue, we have a blue triangle; if not, we have a red triangle $bcdb$.


Frank Plumpton Ramsey (1903–30) and Paul Erdős (1913–96).

Ramsey's theorem can be used to give a simple proof of Schur's result, and other applications were soon apparent. In 1934 a paper on a geometrical theme was written by Paul Erdős and George Szekeres [13]. Erdős stood head and shoulders above any other 20th-century mathematician working in combinatorial areas. Living a life entirely devoted to mathematics he toured the world, living out of a suitcase that contained most of his worldly possessions and sowing the seeds of mathematical ideas wherever he found an 'open' mind that could respond. Altogether he wrote over 1500 research papers, including joint work with more than 500 collaborators throughout the world. He created new areas of research, such as the introduction of probabilistic arguments to establish existence theorems, and asked searching questions that led to deep new insights.

His joint paper with George Szekeres, 'On a combinatorial problem in geometry', appeared in 1935. It discussed a problem attributed to the young Esther Klein (later to become Esther Szekeres):

Given sufficiently many points in the plane, can we always select n of them which form a convex n-sided polygon?

The authors of the paper gave two different proofs of this result; the first, constructed before the authors had come across Ramsey's work, used (and independently proved) Ramsey's theorem.

The Erdős–Szekeres paper improved upon Ramsey's original proof in one important respect. Ramsey theorems assert that something must happen provided that n is 'sufficiently large'. How large? We have seen that, provided six people are present, there must be either three mutual acquaintances or three mutual strangers. How many need to be present to ensure at least a mutual acquaintances or b mutual strangers? The least number needed is denoted by $R(a, b)$, and these numbers are now called *Ramsey numbers*. Thus, as we have seen, $R(3, 3) = 6$. Except for obvious cases such as $R(a, 2) = a$ for all a, very few Ramsey numbers are known; for example, it is known that $R(4, 4) = 18$, but $R(5, 5)$ is unknown. Erdős and Szekeres showed that $C(a + b - 2, a - 1)$ people would suffice, and this bound stood for many years as the best known. Details of the history of these and similar bounds can be found in Winn [45].

Configurations

At the end of the 19th century there was a growing interest in the study of finite geometries. In particular, structures called *plane configurations* n_k were discussed, with n points and n lines and having the properties that, for some k, each line contains k points, each point lies on k lines, and any two points appear on at most one line.

In 1894 Ernst Steinitz, who was to become better known for his work on field theory and polyhedra, was a student at Breslau in the German Empire (now Wrocław in Poland). His dissertation, *Über die Construction der Configurationen* n_3 [40], contained the remarkable result that, if the lines in n_k (which are sets of k points) are written as the columns of a $k \times n$ array, then the elements (points) in each column can be rearranged so that each horizontal row contains each element exactly once. For example, starting from the configuration 8_3

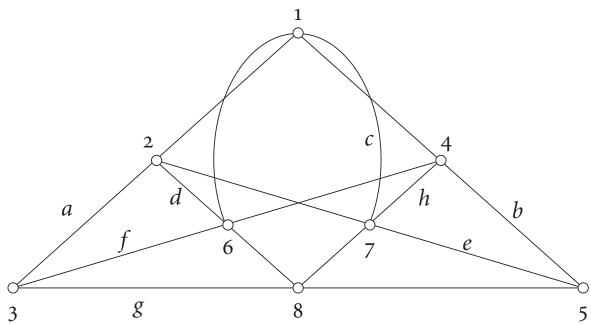
$$\begin{matrix} 1 & 1 & 1 & 2 & 2 & 3 & 3 & 4 \\ 2 & 4 & 6 & 6 & 5 & 4 & 5 & 7 \\ 3 & 5 & 7 & 8 & 7 & 6 & 8 & 8 \end{matrix}$$

we can rearrange the entries in the columns to give

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

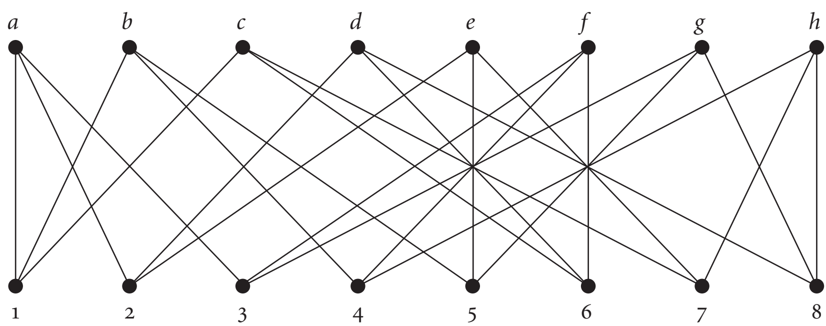
where each row now consists of the numbers $1, 2, \ldots, 8$.

Steinitz used his result to prove that n_3 can be drawn in the plane in such a way that all but one of the lines is straight. One example, the Fano plane 7_3, discussed by Gino Fano [14] in 1892, was depicted in Chapters 10 and 11. The following diagram shows the above configuration 8_3.



The configuration 8_3.

Although Steinitz's result was long neglected (see Gropp [19]), it was essentially rediscovered twenty years later in the language of graph theory. If we label the columns (lines) of the above 8_3 by $a, b, \ldots, h$, then we can represent 8_3 by a graph in which, for example, a is joined to 1, 2, and 3 since line a consists of the points 1, 2, 3. This graph is regular and bipartite: the top vertices have degree 3 since each line has three points, and the bottom vertices have degree 3 because each point lies on three lines.



The configuration 8_3 as a regular bipartite graph.

At the Congress for Mathematical Philosophy in Paris in 1914, Dénes König of the Technische Hochschule in Budapest presented a paper containing the general graph-theoretical result (published later in [28]) that any regular bipartite graph has a 1-factor; this means that, in any such graph with $2n$ vertices, there exist n disjoint edges. In the above 8_3 graph, the eight edges a–1, b–4, c–7, d–2, e–5, f–6, g–3, h–8 form a 1-factor, and give the first row of Steinitz's rearranged array. Since Steinitz did not use the 'plane' property that any two points are joined by at most one line, it follows that König's result is really equivalent to that of Steinitz. In his book *Theorie der endlichen und unendlichen Graphen* (Theory of Finite and Infinite Graphs) [29], the first major book devoted to the theory of graphs, König described how this result relates to many others.

Transversals

What Steinitz did was to choose a different point from each line so as to form the first row of his array. Such a collection of elements, a distinct element chosen from each of a given collection of sets, is now called a *system of distinct representatives* or *transversal*. In 1935 Philip Hall, working at Cambridge, published his famous paper [21] in which he obtained a simple necessary and sufficient condition for a collection of subsets to have a transversal:

The sets $A_1, \ldots, A_m$ possess a transversal $x_1, \ldots, x_m$ with $x_i \in A_i$ for each i, if and only if, for each $k \leq m$, the union of any k of the sets A_i contains at least k elements.

This result gives an easy proof of the Steinitz–König theorem as a special case. It has come to be known as *Hall's theorem*, and the above condition as *Hall's condition*, although it is closely related to results of other mathematicians. For example, at the same time as Hall's paper was being published in the *Journal of the London Mathematical Society*, Maak [32] proved essentially the same result. Maak's work was discussed a decade later by Hermann Weyl, who presented Hall's theorem in terms of friendships between boys and girls [43]; the sets were the lists of girls that the boys were friendly with, and the problem was to marry each boy to a girl who was a friend. This marriage version was popularized by a paper of Halmos and Vaughan [22], who also gave a simpler proof, and the problem is now often called the *marriage problem*.

Historically, Hall's theorem played a key role in the development of combinatorial set theory. It may be equivalent to other results, such as those of König and others, but its formulation is particularly useful in applications. One simple application is due to Marshall Hall, who wrote a major book entitled

Combinatorial Theory and who used the Hall condition in 1945 to prove (see [20]) that any $k \times n$ latin rectangle with $k < n$ can be extended to an $n \times n$ latin square.

Another application of Hall's theorem is to give a simple proof of an earlier result of van der Waerden. In 1910 G. A. Miller [34] had published a result on cosets of a finite group. In its simplest form it states that if a group G is expressed as a disjoint union of n left cosets of a subgroup H, and also as the union of n right cosets of H, then the left and right cosets possess a common transversal. In 1927 van der Waerden [42] showed that this result is not really a group-theoretical result, but a combinatorial one:

If a finite set S with mn elements is partitioned into n sets of size m in two different ways,

$$S = A_1 \cup A_2 \cup \cdots \cup A_n = B_1 \cup B_2 \cup \cdots \cup B_n,$$

then there exist elements $x_1, x_2, \ldots, x_n$ *that act as a transversal for the sets* A_i *and simultaneously for the sets* B_i.

Van der Waerden did not use either König's theorem or Hall's theorem, but gave an independent proof; however, in a note at the end of his paper added at the proof stage, he remarked that his result is equivalent to König's. König deduced van der Waerden's result from the König–Steinitz theorem in his book, whereas Hall in [21] derived it from his own theorem.


Dénes König (1884–1944) and Philip Hall (1904–1982).

It is fascinating to discover how, in the period up to 1940, many results were obtained that are intimately related to those of Hall and König. König himself presented one such result in 1931 to the Budapest Mathematical and Physical Society, of which he was secretary for many years; since another version was presented later in that year by E. Egerváry [9], it is now known as the *König–Egerváry theorem*. Consider, as an example, the $(0,1)$-matrix

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

The smallest number of lines (rows or columns) that together include all the 1s is four (take rows 1 and 4 and columns 2 and 4), while the largest number of 1s that can be chosen with no two in the same row or column is also four (take the 1s in positions $(1,1)$, $(2,2)$, $(3,4)$, and $(4,5)$). The general result is that the *minimum* number of rows and columns that cover all the 1s is always equal to the *maximum* number of 'independent' 1s in the matrix. This is one of many such 'max–min' results; another is Menger's theorem (see Chapter 14).

Dénes König's book [29] remained the definitive exposition of graph theory for thirty years, although it was not translated into English until 1990, and some of the historical footnotes make fascinating reading. Of particular interest are those revealing an apparent degree of hostility between König and Georg Frobenius over a result in the theory of determinants – that, if $\mathbf{A}$ is an $n \times n$ matrix, then every product in the expansion of the determinant det $\mathbf{A}$ is 0 if and only if $\mathbf{A}$ contains a zero $h \times k$ submatrix $\mathbf{B}$ for some h and k with $h + k > n$. In 1912 Frobenius [15] proved this result in what König called 'an extraordinarily complicated way'. König translated it into a graph theory problem (it is essentially equivalent to the Steinitz–König theorem) and sent Frobenius a copy of his simpler proof. Frobenius [16] then published a simpler proof of his own, making no reference to König's proof, but dismissing the use of graphs in determinantal problems as 'a method of little value for the development of the theory of determinants'. This claim was robustly rejected by König in his book.

Sperner's theorem

Around the same time as the above developments were taking place, another fundamental result, which also has a max–min interpretation, was discovered by Emanuel Sperner in Hamburg. A *chain* of subsets of a set X is a sequence $A_1 \subset A_2 \subset \cdots \subset A_m$ of subsets of X, with each subset properly contained in the next. The 'opposite' of a chain is an *antichain*, a collection of subsets none of which contains another. Sperner was asked by the young Otto Schreier (who died tragically at the age of 28, but not before doing fundamental work in combinatorial group theory) how large an antichain of subsets of an n-element set can be. Sperner [39] showed that the largest antichain has size $C(n, \lfloor \frac{1}{2}n \rfloor)$, the number of subsets of size $\lfloor \frac{1}{2}n \rfloor$. This result can be looked on as a max–min result, since the set of all subsets of a set of size n can be partitioned into $C(n, \lfloor \frac{1}{2}n \rfloor)$ chains, but no fewer. So the maximum size of an antichain is equal to the minimum number of chains into which the set of subsets can be partitioned. It was later proved by Dilworth [7] that this is true of any partially ordered set P:

The maximum size of an antichain in P is equal to the minimum number of chains whose union is P.

Sperner's theorem has had applications in many diverse areas. One interesting use of this theorem was to the problem of finding how many of the sums $\varepsilon_1 x_1 + \varepsilon_2 x_2 + \cdots + \varepsilon_n x_n$ (where each ε_i is 1 or -1 and the x_i are real numbers with $|x_i| \geq 1$) can lie in an interval of length 2; this problem had been studied by Littlewood and Offord [31] in 1943. In 1945 Erdős [11] observed that Sperner's theorem can be used to show that the maximum number of such sums is $C(n, \lfloor \frac{1}{2}n \rfloor)$. It was shown later, by Kleitman [26] and Katona [24] independently, that the same bound holds for the number of sums of the same form lying inside a unit circle, when the x_i are two-dimensional vectors of length at least 1; again the proof depends upon Sperner's theorem. Then, in 1970, by an ingenious partition of the set of sums into blocks of a particular form that imitated chains of subsets, Kleitman [27] extended the result to vectors in an arbitrary number of dimensions. This completely solved what had become known as the *Littlewood–Offord problem*.

Sperner's theorem plays a key role in modern combinatorial set theory. Its generalizations and ramifications have led to a whole new area called *Sperner theory*. Further details can be found in the books by Anderson [1] and Engel [10].

The Erdős–Ko–Rado and Kruskal–Katona theorems

During the 1930s many mathematicians, including Artin, Courant, Noether, von Neumann, and Weil, left continental Europe, seeking refuge from the growing Nazi threat. Also among these was Richard Rado, who left Berlin in 1933 and went to Cambridge to work under G. H. Hardy; thereafter he held university positions in Sheffield, London, and Reading. It was at Cambridge in 1934 that Rado first met Erdős, who had left Hungary for a four-year postdoctoral fellowship in Manchester. Along with the Chinese mathematician Chao Ko, who was also at Manchester and studying for a PhD degree under L. J. Mordell, Erdős and Rado discovered the far-reaching answer to the following problem:

If $k \leq \frac{1}{2}n$, how many k-element subsets $A_1, A_2, \ldots, A_m$ of an n-element set X can be found, such that $A_i \cap A_j \neq \emptyset$ for all i, j?

Certainly, we could select one of the elements x of X and take all the k-element subsets containing x; this would give $C(n-1, k-1)$ intersecting sets. By a fairly technical method they showed in 1938 that this is indeed the largest possible value of m. This result, known as the *Erdős–Ko–Rado theorem* [12], lay unpublished until 1961, an unusually long publication delay. Erdős wrote that one of the reasons for the delay was that there was relatively little interest in the subject at the time. However, the 1960s was a decade that saw a great increase of interest in such matters, and their result was soon to be extended and generalized in many ways. For example, Hilton and Milner [23] showed in 1967 that if the intersecting k-element subsets do not all have an element in common, then

$$m \leq C(n-1, k-1) - C(n-k-1, k-1) + 1.$$

Another important development was the deduction of the Erdős–Ko–Rado theorem from the Kruskal–Katona theorem which we describe below.

The power of modern computers to search through vast amounts of data has led to considerations of methods for storing lists or sequences:

Is there a good way of listing all k-element subsets of $\{1, 2, \ldots, n\}$*?*

One method is to use the *lexicographic ordering*: list the subsets in numerical order; for example, the 3-element subsets of $\{1, 2, 3, 4, 5\}$ in lexicographic order are

$$\{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,3,4\}, \{1,3,5\},$$

$$\{1,4,5\}, \{2,3,4\}, \{2,3,5\}, \{2,4,5\}, \{3,4,5\}.$$

Here we place set A before set B if the *smallest* element that is in one of A and B, but not both, lies in A.

Although this seems a very natural ordering, there is a more important one, called the *squashed* ordering. Here we place set A before set B if the *largest* element in one or the other, but not in both, lies in B. This ordering has the effect of squeezing together early in the ordered list all those sets that do not have a 'large' element. The 3-element subsets of $\{1,2,3,4,5\}$ in squashed order are

$$\{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \{1,2,5\},$$

$$\{1,3,5\}, \{2,3,5\}, \{1,4,5\}, \{2,4,5\}, \{3,4,5\}.$$

This ordering may seem more natural once it is realized that the squashed ordering of the sets corresponds to the lexicographic ordering of the binary sequences of length 5 containing three 1s:

00111, 01011, 01101, 01110, 10011, 10101, 10110, 11001, 11010, 11100;

for example, the set $\{1,2,4\}$ corresponds to the binary sequence 01011, since the 1s appear in the 1st, 2nd, and 4th positions from the right.

Suppose that we are now given four subsets of $\{1,2,3,4,5\}$ of size 3. How few subsets of size 2 can be contained in them? If we take four subsets at random, say

$$\{1,2,4\}, \{2,3,5\}, \{1,3,4\}, \{2,3,4\},$$

then they contain eight 2-element subsets – namely,

$$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{2,5\}, \{3,4\}, \{3,5\}.$$

The first four subsets in the lexicographic ordering also contain eight 2-element subsets, but the first four in the squashed ordering contain only six. This illustrates a general result, stated without proof by Schützenberger [38] in 1959, but proved independently by Kruskal [30] and Katona [25] a few years later, that:

if r k-element subsets of $\{1,2,\ldots,n\}$ are to be chosen so as to minimize the number of $(k-1)$-element subsets contained in them, then the first r sets in the squashed ordering should be selected.

This result, known as the *Kruskal–Katona theorem*, is reminiscent of an earlier theorem proved in 1927 by Francis Sowerby Macaulay, a schoolteacher who included G. N. Watson and J. E. Littlewood among his pupils. He wrote fourteen research papers, mainly on algebraic geometry and polynomial ideals, and was elected a Fellow of the Royal Society in 1928, the year after his paper [33] on a combinatorial topic appeared in the *Journal of the London Mathematical Society*. In that paper Macaulay proved that if the set of vectors of dimension n with non-negative integer components is made into a partially ordered set by defining

$$(a_1, a_2, \ldots, a_n) \leq (b_1, b_2, \ldots, b_n), \text{ whenever } a_i \leq b_i \text{ for all } i,$$

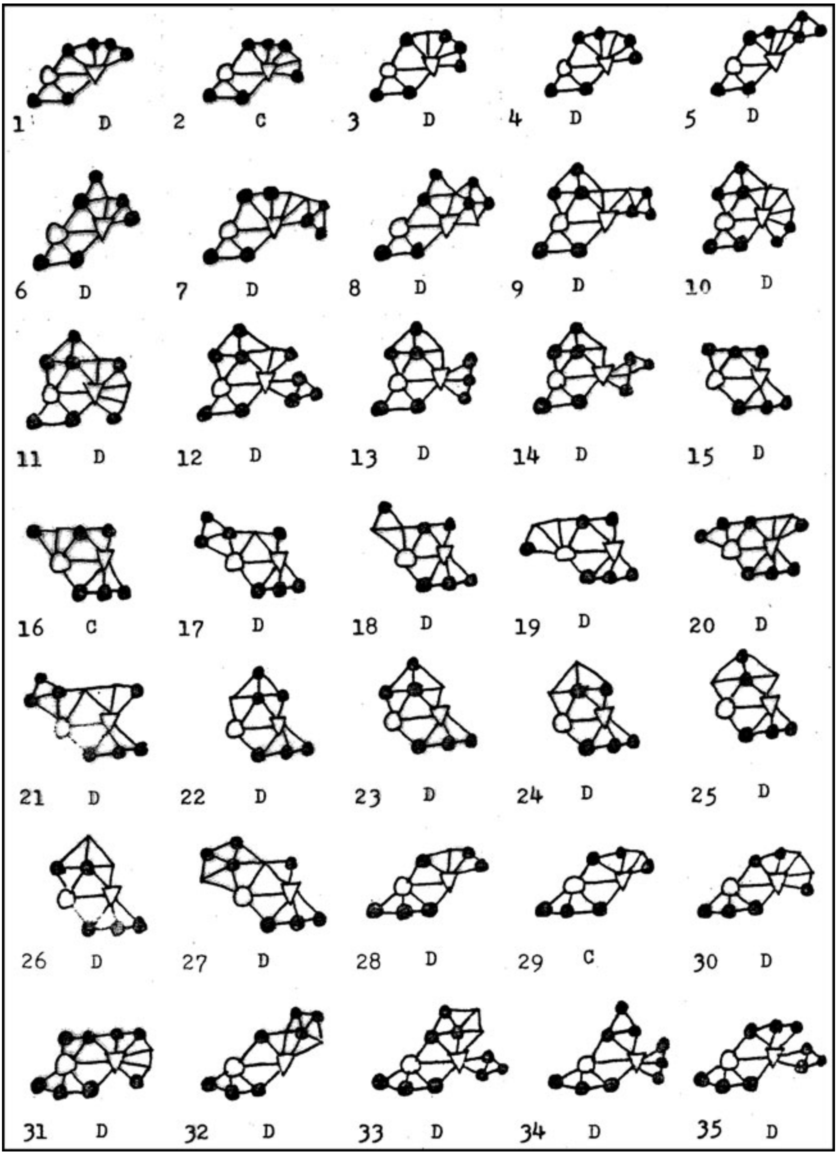
and if each vector $(a_1, a_2, \ldots, a_n)$ is given the rank $a_1 + a_2 + \cdots + a_n$, then we can choose r vectors of rank k that cover the smallest possible number of vectors of rank $k - 1$ by taking the first r vectors of rank k in the lexicographic ordering. Since then, these ideas have led to the modern study of *Macaulay posets*, partially ordered sets in which there is an ordering that gives a minimization result as above. One of the most important results in this area is the Clements–Lindström theorem [3], which states that the partially ordered set of divisors of an integer is a Macaulay poset. This result includes both the Kruskal–Katona theorem and Macaulay's theorem as special cases; further details can be found in Engel [10].

References

1. I. Anderson, *Combinatorics of Finite Sets*, Clarendon Press, Oxford (1987) (reprinted by Dover, 2002).
2. V. Brun, Le crible d'Eratosthène et le théorème de Goldbach, *Skr. Vidensk. Selsk. Christ.* 1 (1920).
3. G. F. Clements and B. Lindström, A generalisation of a combinatorial theorem of Macaulay, *J. Combin. Theory* 7 (1969), 230–8.
4. A. de Moivre, *The Doctrine of Chances*, Pearson (1718).
5. P. R. de Montmort, *Essay d'Analyse sur les Jeux de Hazard*, Paris (1708).
6. L. E. Dickson, On the congruence $x^n + y^n + z^n \equiv 0 \pmod{p}$, *J. Reine Angew. Math.* 135 (1909), 134–41.
7. R. P. Dilworth, A decomposition theorem for partially ordered sets, *Ann. Math.* 51 (1950), 161–5.
8. P. G. L. Dirichlet, *Vorlesungen über Zahlentheorie*, F. Vieweg (1879).
9. E. Egerváry, Mátrixok kombinatorikus tulajdonságairól, *Mat. Fiz. Lapok* 38 (1931), 16–28.

10. K. Engel, *Sperner Theory*, Cambridge University Press, 1997.
11. P. Erdős, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* 51 (1945), 898–902.
12. P. Erdős, Chao Ko, and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math.* 12 (1961), 313–20.
13. P. Erdős and G. Szekeres, On a combinatorial problem in geometry, *Compositio Math.* 2 (1935), 463–70.
14. G. Fano, Sui postulari fondamenti della geometria proiettiva, *Giornale di Matematiche* 30 (1892), 106–32.
15. G. Frobenius, Über Matrizen aus nicht negativen Elementen, *Sitzungsber. Preuss. Akad.* (1912), 456–77.
16. G. Frobenius, Über zerlegbare Determinanten, *Sitzber. König Preuss. Akad. Wiss.* 18 (1917), 274–7.
17. C. F. Gauss, *Disquisitiones Arithmeticae*, G. Fleischner (1801); English transl. by A. A. Clarke, Yale University Press (1966).
18. A. M. Gleason, R. E. Greenwood, and L. M. Kelly (eds), *The William Lowell Putnam Mathematical Competition Problems and Solutions 1938–1964*, Mathematical Association of America (1980).
19. H. Gropp, On combinatorial papers of König and Steinitz, *Acta Appl. Math.* 52 (1998), 271–6.
20. M. Hall, An existence theorem for Latin squares, *Bull. Amer. Math. Soc.* 51 (1945), 387–8.
21. P. Hall, On representations of subsets, *J. London Math. Soc.* 10 (1935), 26–30.
22. P. R. Halmos and H. E. Vaughan, The marriage problem, *Amer. J. Math.* 72 (1950), 214–15.
23. A. J. W. Hilton and E. C. Milner, Some intersection theorems for systems of finite sets, *Quart. J. Math. Oxford (2)* 18 (1867), 369–84.
24. G. O. H. Katona, On a conjecture of Erdős and a strong form of Sperner's theorem, *Studia Sci. Math. Hungar.* 1 (1966), 59–63.
25. G. O. H. Katona, A theorem of finite sets, *Theory of Graphs* (ed. P. Erdős and G. O. H. Katona), Akademia Kaido and Academic Press (1968), 187–207.
26. D. J. Kleitman, On a lemma of Littlewood and Offord on the distribution of certain sums, *Math. Z.* 90 (1965), 251–9.
27. D. J. Kleitman, On a lemma of Littlewood and Offord on the distribution of linear combinations of vectors, *Adv. Math.* 5 (1970), 1–3.
28. D. König, Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, *Math. Ann.* 77 (1916), 453–65.
29. D. König, *Theorie der endlichen und unendlichen Graphen*, Akademische Verlag (1936); English transl. by R. McCoart, *Theory of Finite and Infinite Graphs*, Birkhäuser (1990).
30. J. B. Kruskal, The number of simplices in a complex, *Mathematical Optimization Techniques* (ed. R. Bellman), University of California Press (1963), 251–78.
31. J. E. Littlewood and A. C. Offord, On the number of real roots of a random algebraic equation III, *Mat. Sbornik* 12 (1943), 277–86.

32. W. Maak, Eine neue Definition der fastperiodischen Funktionen, *Abh. Math. Sem. Hamburg. Univ.* 11 (1936), 240–4.
33. F. S. Macaulay, Some properties of enumeration in the theory of modular systems, *Proc. London Math. Soc. (2)* 26 (1927), 531–55.
34. G. A. Miller, On a method due to Galois, *Quart. J. Pure Appl. Math.* 41 (1910), 382–4.
35. K. Ollerenshaw and D. Brée, *Most-Perfect Pandiagonal Magic Squares*, Institute of Mathematics and Its Applications, Southend-on-Sea (1998).
36. F. P. Ramsay, On a problem in formal logic, *Proc. London Math. Soc. (2)* 30 (1930), 264–86.
37. I. Schur, Über die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *Jber. Deutsch Math. Verein.* 25 (1916), 114–16.
38. M. Schützenberger, A characteristic property of certain polynomials of E. H. Moore and C. E. Shannon, *Res. Lab. Electron. Quart. Rep., MIT* 55 (1959), 117–18.
39. E. Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* 27 (1928), 544–8.
40. E. Steinitz, *Über die Construction der Configurationen* n_3, PhD thesis, University of Breslau (1894).
41. B. L. van der Waerden, Beweis einer baudetschen Vermutung, *Nieuw Arch. Wisk.* 15 (1927), 212–16.
42. B. L. van der Waerden, Ein Satz über Klasseneinteilungen von endlichen Mengen, *Abh. Math. Sem. Hamburg Univ.* 5 (1927), 185–8.
43. H. Weyl, Almost periodic invariant vector sets in a metric vector space, *Amer. J. Math.* 71 (1949), 178–205.
44. H. Whitney, A logical expansion in mathematics, *Bull. Amer. Math. Soc.* 38 (1932), 572–9.
45. J. A. Winn, *Asymptotic Bounds for Classical Ramsay Numbers*, Polygonal (1988).



A page of Appel and Haken's reducible configurations.

CHAPTER 14

Modern graph theory

LOWELL BEINEKE AND ROBIN WILSON

During the first half of the 20th century many classic theorems about graphs were discovered, but it was not until the second half of the century that graph theory emerged as an important field in its own right. This chapter develops themes arising from the four-colour problem, before focusing on three specific subject areas – the factorization of graphs, connectivity, and graph algorithms [1].

Planar graphs

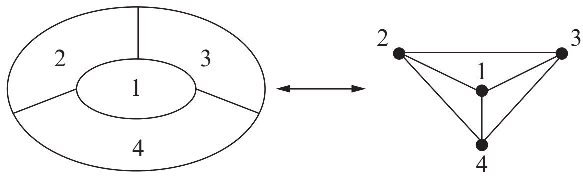
Like many other aspects of graph theory, the origins of the study of planar graphs can be found in recreational puzzles. One such poser was given by August Möbius in his lectures around the year 1840 (see [5, pp. 115–16]):

> There was once a king with five sons. In his will he stated that after his death the sons should divide the kingdom into five regions, so that the boundary of each region should have a frontier line in common with each of the other four regions. Can the terms of the will be satisfied?

This question asks whether it is possible to find five mutually neighbouring regions in the plane. We can turn this into a graph theory problem by 'dualizing' it, replacing regions by capital cities and frontier lines by connecting roads, as follows:

There was once a king with five sons. In his will he stated that after his death the sons should build non-intersecting roads joining the five capital cities in the regions of his kingdom. Can the terms of the will be satisfied?

If there had been only four sons, then both problems could have been easily solved; the following figure illustrates the solutions and a dual connection between them.

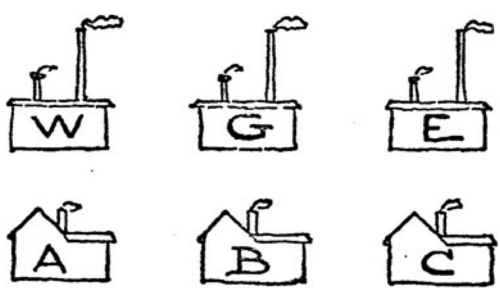


Defining the *complete graph* K_n to be the graph obtained by joining n vertices in pairs, we see that K_4 is *planar* – that is, it can be drawn without any edges crossing, as illustrated above. However, a little experimentation, or use of Euler's polyhedron formula (see Chapter 8), shows that in any drawing of K_5 some edges must cross. Thus:

The graph K_5 *is not planar*.

So in Möbius's two puzzles the terms of the will cannot be satisfied.

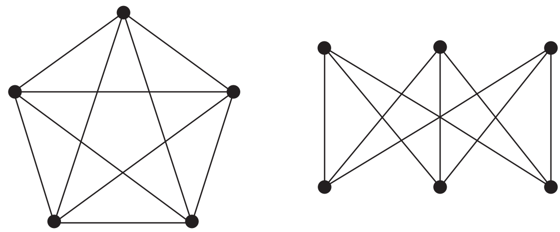
A related problem is the *'gas–water–electricity' problem*. Its origins are obscure, but in 1913 Henry Dudeney [14] presented the problem as follows, describing it as 'as old as the hills':



The puzzle is to lay on water, gas, and electricity, from W, G, and E, to each of the three houses, A, B, and C, without any pipe crossing another. Take your pencil and draw lines showing how this should be done. You will soon find yourself in difficulties . . . .

This problem also has no solution, although Dudeney claimed to have solved it by running a pipe through one of the houses. Defining the *complete bipartite graph* $K_{r,s}$ to be the graph obtained by joining each of r vertices to each of s other vertices, then our result is:

The graph $K_{3,3}$ *is not planar*.



The 'Kuratowski graphs' K_5 and $K_{3,3}$.

In 1930 Kazimierz Kuratowski [29] published the surprising result that *these two graphs are the only 'basic' non-planar graphs*, in the sense that every non-planar graph must contain at least one of them, a result obtained independently by O. Frink and P. A. Smith.

For some time, mathematicians tried to find characterizations of planar graphs that depend on combinatorial, rather than geometrical, considerations. The clue to doing so turned out to be through duality; note that only planar graphs, such as K_4, have geometrical duals, as illustrated above. In 1931 Hassler Whitney [51] formulated an abstract definition of duality that is purely combinatorial (involving the cycles and cutsets of two graphs) and which agrees with the geometrical definition of a dual graph when the graph is planar. He then proved the following for this abstract form of dual:

A graph is planar if and only if it has an abstract dual.

Extending these ideas eventually led Whitney to the concept of a *matroid*, which generalizes the ideas of 'independence' in both graphs and vector spaces [52]; in particular, the dual of a matroid is a natural concept that extends and clarifies the duality of planar graphs. Interest in matroids took some time to develop, but in the 1950s W. T. Tutte [46] obtained a Kuratowski-type condition for a matroid to be one that arises from a graph.

The four-colour theorem

Following Percy Heawood's 1890 bombshell [21] (see Chapter 8), potential solvers of the four-colour problem were back at square one. Over the next eighty years, the slow climb back involved two main ideas, those of an *unavoidable set of configurations* and a *reducible configuration*.

In his 1879 paper, Kempe had shown that every map necessarily contains a digon, triangle, quadrilateral, or pentagon. Since at least one of these configurations must appear, we call such a set of configurations an *unavoidable set*. He also showed that if a map contains a digon, a triangle, or a quadrilateral, then any colouring of the rest of the map can be extended to include this configuration. Any configuration of countries for which this is true is called *reducible*. Note that no reducible configuration can appear in a minimal counter-example to the four-colour theorem.

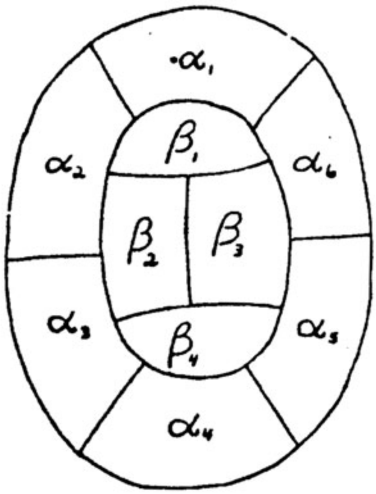
Where Kempe's proof broke down is that he failed to prove that a pentagon is reducible, and the search began for configurations that might replace the pentagon in the unavoidable set. In 1904 Paul Wernicke [50] proved that it can be replaced by a pair of adjacent pentagons and a pentagon adjacent to a hexagon, thereby obtaining a more complicated unavoidable set that could then be tested for reducibility. Later, in 1922, Philip Franklin [16] proved that every cubic map containing no digons, triangles, or quadrilaterals must have at least twelve pentagons, and must include at least one of the following:

- a pentagon adjacent to two other pentagons;
- a pentagon adjacent to a pentagon and a hexagon;
- a pentagon adjacent to two hexagons.

Using this unavoidable set he proved the four-colour theorem for maps with up to twenty-five countries. Unavoidable sets were also given by C. N. Reynolds, Henri Lebesgue (mainly known for his work on the integral calculus), and others, and over the years the four-colour theorem came to be proved for larger and larger maps.

Meanwhile, the search was on for reducible configurations other than the digon, triangle, and quadrilateral. G. D. Birkhoff [7], who learned of the four-colour problem from Oswald Veblen while studying at Princeton

University, showed that various other configurations, such as the 'Birkhoff diamond' $\beta_1, \beta_2, \beta_3$, and β_4 of four adjacent pentagons, are reducible.



The Birkhoff diamond.

This two-pronged attack of constructing unavoidable sets and proving configurations to be reducible would eventually prove successful. On the one hand one could replace the pentagon by more and more complicated unavoidable sets, and on the other hand one could try to obtain larger and larger lists of reducible configurations. The ultimate aim was to find *an unavoidable set of reducible configurations*, since every map would have to contain at least one such configuration, and whichever it was, any colouring of the rest of the map could then be extended to the configuration. Alternatively, since every map must contain one of these reducible configurations, and since no reducible configuration can appear in a counter-example to the four-colour theorem, there can be no such counter-example.

Around 1970 Heinrich Heesch [22] presented arguments that indicated that a finite unavoidable set of reducible configurations existed, and that the number of such configurations would not exceed 9000. In addition, he developed a technique for constructing unavoidable sets, later called the *discharging method*, and noticed that there are certain features of a map that seem to prevent a configuration from being reducible.

Kenneth Appel (b.1932) and Wolfgang Haken (b.1928).

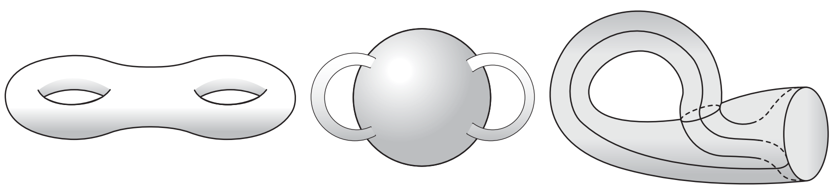
These ideas were developed by Kenneth Appel and Wolfgang Haken, who spent several years designing computer programs that would help in the search for unavoidable configurations and assist in testing for reducibility. Unlike other investigators, who created large numbers of reducible configurations and then tried to package them into unavoidable sets, Appel and Haken's approach was to construct unavoidable sets of 'likely-to-be-reducible' configurations and then check them for reducibility, modifying the set as necessary – this approach saved much time and effort. After some 1200 hours of computer time, they eventually produced an unavoidable set of 1936 reducible configurations (later reduced to 1482), thereby completing the proof of the four-colour theorem (see [2] and [3]). Indeed, since their approach yielded many thousands of such unavoidable sets, they had thousands of proofs of the four-colour theorem, so that if any individual configuration were subsequently to be proved irreducible, this would not invalidate their work. For further details of their proof, see [53].

Since then, the technical details of the proof have been simplified somewhat, mainly by Robertson, Sanders, Seymour, and Thomas (see [39] and [42]), and the configurations have been checked on other computers, but no easily verifiable proof has yet been found. Because of this, and because Appel and Haken's work raised interesting philosophical questions about the nature of mathematical proof, the mathematical world was slow to acclaim their magnificent achievement (see [53, Ch. 11]).

Higher surfaces

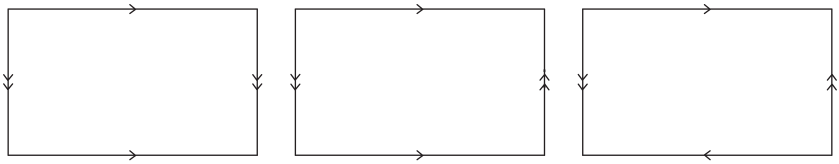
Our earlier discussion on the colouring of maps drawn on the plane applies equally well to maps drawn on a sphere, and this leads us to ask how many colours are needed for maps drawn on other surfaces. Indeed, in the 1890 paper in which he demolished Kempe's 'proof' (see Chapter 8), Heawood [21] raised this very question, proving that seven colours are sufficient for all maps drawn on a torus and that for some maps seven colours are actually needed.

Surfaces are of two types. On the one hand are the *orientable surfaces*, which can be thought of as spheres with a number of handles added; for example, the double torus is a sphere with two handles added. On the other hand are the *non-orientable surfaces*, with a 'twist' (or 'cross-cap') in them. These include the *projective plane* and the *Klein bottle*.



Double torus, sphere with two handles, and Klein bottle.

Any surface can be obtained by taking a polygon and identifying some of its edges. For example, when we identify opposite edges of a rectangle we obtain a torus, but if we give one pair of opposite edges a twist before identifying them, we obtain a Klein bottle. For the projective plane, we give both pairs of opposite edges a twist.



Identifying edges to obtain a torus, a Klein bottle, and a projective plane.

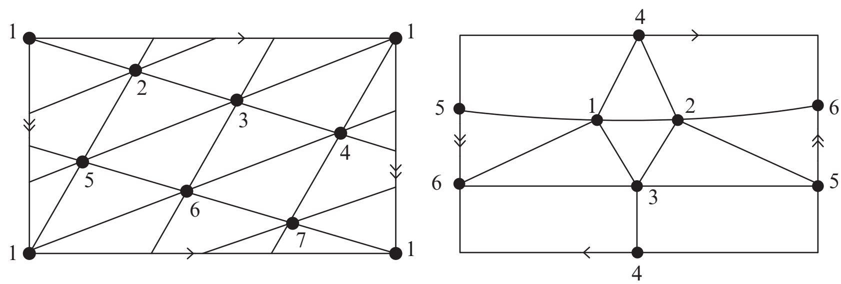
Using Lhuilier's result that the Euler characteristic of an orientable surface with g handles is $2 - 2g$ (see Chapter 8), Heawood showed that any map drawn on such a surface can be coloured with $\lfloor \frac{1}{2}\left(7 + \sqrt{1 + 48g}\right) \rfloor$ colours when $g \geq 1$; this reduces to 7 for a torus ($g = 1$) and to 8 for a double torus ($g = 2$). Unfortunately, except for the torus, he omitted to prove that there are maps that actually require this number of colours, claiming that 'there are generally contacts enough and to spare'.

In the following year, Lothar Heffter [23] pointed out the omission, and provided the necessary constructions for certain specific values of g. Shortly after, Heinrich Tietze [43] raised the corresponding problem for maps drawn on *non-orientable* surfaces. For a non-orientable surface with g cross-caps, the Euler characteristic is $2 - g$, and the appropriate upper bound on the number of colours becomes $\lfloor \frac{1}{2}\left(7 + \sqrt{1 + 24g}\right) \rfloor$. It took many years to prove that in both the orientable and the non-orientable cases these upper bounds can be achieved for all remaining values, except for $g = 2$, the Klein bottle, in the non-orientable case. The proof eventually came after a long and difficult struggle involving many people – most notably, Gerhard Ringel and J. W. T. Youngs.

The main problem that Ringel and Youngs solved involved the (*orientable*) *genus* of a graph, the least number of handles that must be added to a sphere for the graph to have a drawing with no edges crossing. For example, the complete graph K_5 has genus 1, since it can be drawn on a torus, but not on a sphere; in fact, the same is true of K_6 and K_7. A key connection between this problem and the four-colour problem is that if the genus of K_n is g_n, then there are some maps on the surface with g_n handles that require n colours. In 1968 Ringel and Youngs showed that the genus of K_n is $\lceil \frac{1}{12}(n-3)(n-4) \rceil$, and from this and the above observation it follows that, for $g \geq 1$, the surface with g handles has maps on it that actually require $\lfloor \frac{1}{2}\left(7 + \sqrt{1 + 48g}\right) \rfloor$ colours.

The Ringel–Youngs proof involved the ingenious use of a related 'electrical current graph' and split into no fewer than twelve separate arguments, depending on the remainder when n is divided by 12. Some of these cases turned out to be particularly intransigent; for details, see the book by Ringel [37]. Intriguingly, a few particular values of n that needed special treatment proved to be rather difficult, and were eventually sorted out by a professor of French literature (see [32])!

Meanwhile, corresponding results had also been obtained for non-orientable surfaces. For such a surface with g cross-caps, Kagno [25] proved in 1935 that there are maps that require $\lfloor \frac{1}{2}\left(7+\sqrt{1+24g}\,\right)\rfloor$ colours when $g = 3$, 4, or 6, but in the previous year Franklin [17] had shown that for the Klein bottle, where $g = 2$, the correct number of colours turns out to be 6, rather than the value 7 given by this formula. In 1952 Ringel obtained the complete solution, proving that there are maps that require this number of colours, with the single exception of maps on the Klein bottle (see [37]).



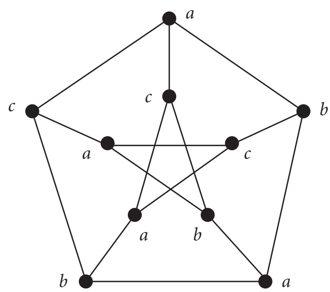
K_7 drawn on the torus and K_6 drawn on the projective plane.

We saw earlier that a graph can be drawn in the plane or sphere if it does not contain either of the 'forbidden subgraphs' K_5 and $K_{3,3}$ (Kuratowski's theorem). The analogous question for higher surfaces, as to whether there is a finite set of 'forbidden subgraphs' for each one, remained elusive for a long time, although in 1979 Glover *et al.* [18] managed to obtain a set of 103 forbidden subgraphs for the projective plane.

Eventually, in a remarkable series of papers in the 1980s, Neil Robertson and Paul Seymour achieved a major result on 'graph minors', one of the most remarkable breakthroughs in the whole of graph theory. One of its consequences is that, for each such surface (orientable or non-orientable), there is indeed a *finite* set of forbidden subgraphs; however, unlike the situation for the sphere, the number of such subgraphs may be very large – even for the torus, there are many hundreds of them. A survey of this topic can be found in Robertson and Seymour [38].

Other colouring problems

As stated, the four-colour map problem does not appear to be a problem in graph theory. However, as Kempe pointed out in his 1879 paper, the map problem can be dualized to give a problem on the colouring of vertices; in this formulation, we are required to colour the vertices of a planar graph with four colours in such a way that any two vertices joined by an edge are coloured differently; this reformulation was the version in which Appel and Haken's solution was presented.

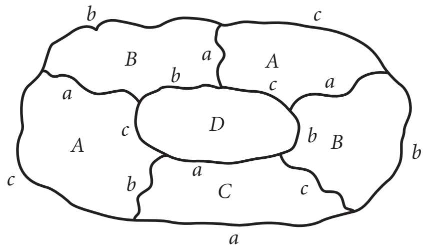


A graph whose vertices are coloured with three colours.

More generally, one can ask for the *chromatic number* of any given graph, the smallest number of colours needed to colour its vertices in such a way that any two vertices joined by an edge are coloured differently. This idea developed a life of its own in the 1930s, mainly through the work of Hassler Whitney, who wrote his PhD thesis on the colouring of graphs (see [50]). In particular, Whitney developed for graphs an idea that Birkhoff [6] had introduced for maps; this is the *chromatic polynomial*, which gives the number of possible colourings as a polynomial function of the number of colours available. Such polynomials can be usefully studied in their own right, as was done to good effect by Birkhoff and Lewis [8], Tutte, and others. Subsequently, R. L. Brooks [10] obtained a useful upper bound on the number of colours required to colour the vertices of a graph, in terms of the largest vertex degree in the graph:

For a connected graph that is not a complete graph or a cycle of odd length, the chromatic number cannot exceed the maximum vertex degree.

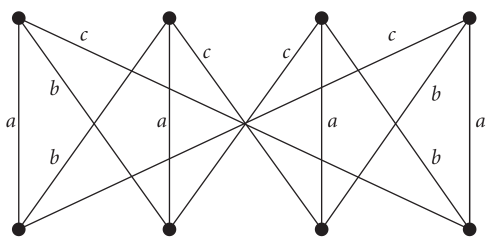
In 1880, while trying to prove the four-colour theorem, P. G. Tait [41] proved that the countries of a cubic map can be coloured with four colours if and only if its boundary edges can be coloured with three colours so that the edges meeting at any point are all coloured differently.



A four-coloured cubic map whose edges are coloured with three colours.

More generally, one can ask for the *chromatic index* of a given graph, the smallest number of colours needed to colour its edges in such a way that any two edges that meet at a vertex are coloured differently. It is clear that if a graph contains a vertex of degree k, then we need at least k colours to colour the edges. It follows that the number of colours needed to colour all the edges of the graph cannot be less than the maximum vertex degree in the graph.

A graph is *bipartite* if its set of vertices can be split into two sets in such a way that each edge joins a vertex in one set to a vertex in the other. In 1916 Dénes König [27] proved that if a bipartite graph has largest vertex degree k, then its edges can be coloured with just k colours; for example, the edges of the bipartite graph below require three colours.



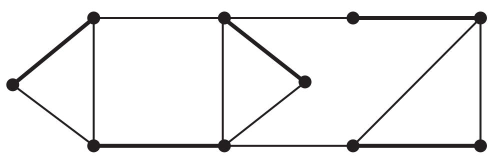
In 1949 Claude Shannon [40] reformulated a problem involving electrical relays as an edge-colouring graph problem and established an upper bound for the number of colours required. Then, in a fundamental paper, V. G. Vizing [47] proved that to colour the edges of any graph with largest vertex degree k, $k + 1$ colours are always sufficient. This led to the still-unsolved *classification problem* of determining which graphs need only k colours and which ones need $k + 1$ colours.

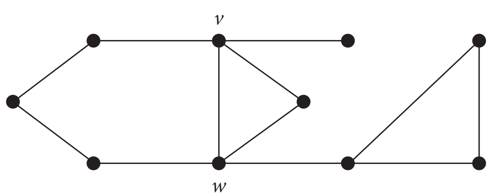
The study of colouring problems of various types blossomed throughout the 1970s and 1980s, and continues to flourish. Further information about these developments can be found in the book by Jensen and Toft [24].

Matchings and factorization

An important problem in graph theory is to determine whether one can pair up the vertices of a given graph in such a way that the vertices in each pair are joined by an edge – in other words, can one find a set of disjoint edges meeting each vertex? Such questions arise, for example, when the vertices represent people who are required to work in pairs, and the edges indicate a willingness to work together.

The set of edges in such a pairing is called a *perfect matching*. For example, the first of the graphs below has a perfect matching, shown with heavy edges. However, the second graph has no perfect matching – for, if the vertices v and w are removed (with their incident edges), then what remains has four pieces, each with an odd number of vertices; in a perfect matching, v and w would have to be paired with vertices from all four pieces, contradicting the definition.





This idea lies at the basis of a result of Tutte [44], published in 1947, that asserts that a graph has a perfect matching unless there are vertices like the above pair v and w:

A graph has a perfect matching if and only if it has no set of k vertices whose removal leaves a graph in which more than k of the connected pieces that remain have an odd number of vertices.

Note that Tutte's theorem does not tell us how to find a perfect matching in a graph (if one exists); in fact, showing that there is one can take a lot of checking.

Another type of matching problem is the *assignment problem*:

A firm has several positions to be filled, and there are several applicants, each of whom is qualified for certain of the positions. When can all of the openings be filled?

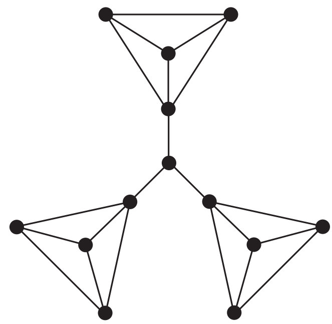
This situation can be modelled with a bipartite graph, with the positions as one set of vertices, the applicants as the other set, and with edges joining the positions to qualified applicants. A solution to the problem is then a set of disjoint edges that covers all of the openings. In set theory terms the solution is given by Hall's theorem (see Chapter 13); for us, this says that all the jobs can be filled if and only if each set of positions has at least as many applicants as the number of positions.

When a solution does not exist for any of these matching problems, it is natural to ask how close one can get, and how one can find an optimal solution. A brief discussion of such questions appears in the last section of this chapter.

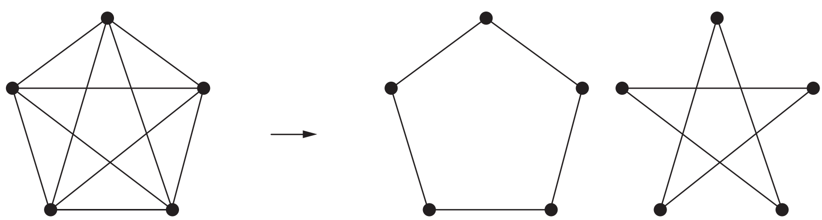
A perfect matching in a graph is often called a 1-*factor*. Perhaps surprisingly, the study of 1-factors did not begin with the matching question, but originated with Tait's 1880 result, mentioned earlier, that a 4-colouring of the countries of a cubic map corresponds to a 3-colouring of the edges in which the colours meeting at each point are all different. In such a 3-colouring the edges of each colour constitute a 1-factor, and taken together the three 1-factors give a 1-*factorization* of the graph. More generally, a graph is *k-regular* if each of its vertices has degree k, and such a graph has a 1-factorization, or is 1-*factorizable*, if its set of edges can be split into k 1-factors.

The first mathematician to study factorizations systematically was Julius Petersen, who in 1891 wrote a fundamental paper [35] on the factorization of regular graphs, arising from a problem in the theory of invariants. For a regular graph of odd degree, the primary question is whether it has a 1-factor and, if so, whether it is 1-factorizable. As shown earlier, Tait's graph has a 1-factorization. However, the following 3-regular graph has no 1-factor: note, however, that it has three 'cut-edges' – if any one of them is removed, then the remaining graph has two pieces. Petersen proved that:

Every 3-regular graph with at most one cut-edge has a 1-factor.



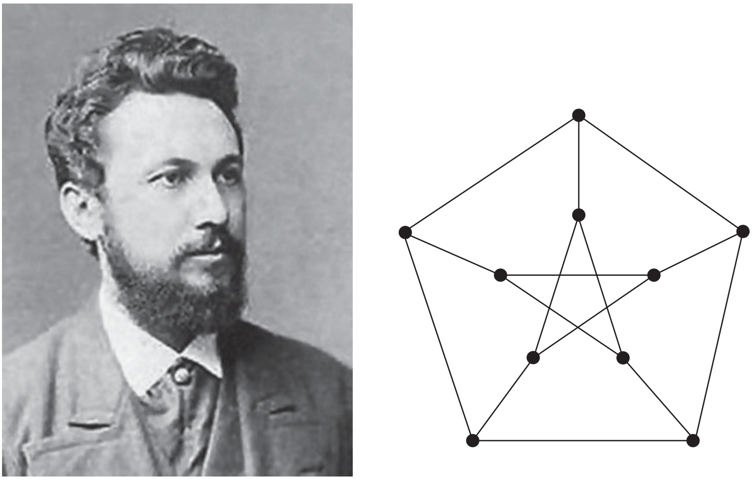
More generally, an *r-factor* in a graph is a subgraph in which each vertex has degree r. After 1-factors, by far the most interesting ones are the 2-factors, collections of cycles with the property that each vertex of the graph lies in just one cycle. A graph is *r-factorizable* if its set of edges can be split up into r-factors; for example, the complete graph K_5, which is 4-regular, is 2-factorizable, since its set of edges can be split into two 2-factors (the outer pentagon and the inner pentagram).



Petersen realized that the factorization of regular graphs of even degree is simpler than for those of odd degree, and he proved that:

Every regular graph of even degree is 2-factorizable.

A few years later, he wrote a short note [36] in which he presented a cubic graph that is not 1-factorizable but can be split into a 2-factor (the pentagon and pentagram) and a 1-factor (the spokes joining them). It is now called the *Petersen graph*.

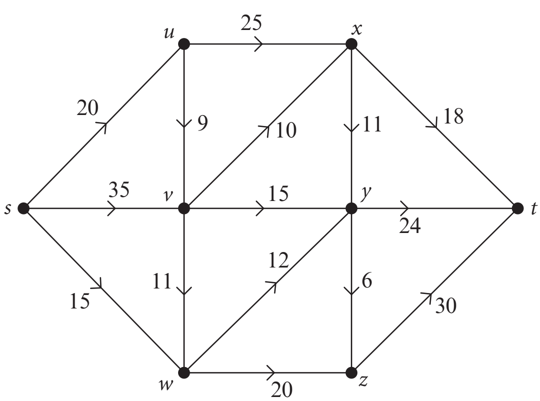


Julius Petersen (1839–1910) and the Petersen graph.

Many years later, following his result on the existence in a graph of a 1-factor (or perfect matching), Tutte [45] proved a corresponding theorem that tells us when a graph has an r-factor.

Connectivity

Another application of graphs lies in their modelling of flows in networks. In this setting, there is a collection of locations with certain pairs connected directly by roads, which may be in just one direction, and each road has a certain capacity for the amount of goods that can be transported along it; the figure below gives an example of such a 'capacitated network'.

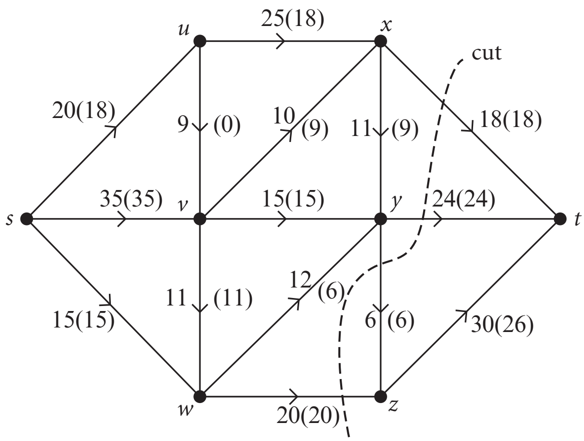


Two vertices are specified as the start *s* and terminus *t*, and the objective is to find *flows* from *s* to *t*, wherein the amount flowing along each arc does not exceed the capacity and (except for *s* and *t*) the amount coming into each vertex equals the amount going out. The total amount flowing from *s* to *t* is called the *value* of the flow, and we wish to find a flow of maximum value.

A *cut* in a network is a collection of arcs whose removal leaves no directed paths from *s* to *t*. Its *capacity* is the sum of the weights on its arcs. In 1956 Ford and Fulkerson [15] proved the renowned *max-flow–min-cut theorem*:

In any network, the maximum value of a flow equals the minimum capacity of a cut.

For the above network, this theorem is illustrated below, where the numbers in brackets give a flow from *s* to *t* of value 68, and the arcs *xt*, *yt*, *yz*, and *wz* form a cut of capacity 68.

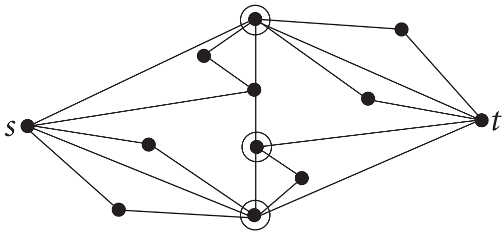


In fact, such considerations can be traced back to the 1920s, to work by Karl Menger and others on the *connectivity* (the degree of connectedness) of graphs. If, in a connected graph, *s* and *t* are two vertices that are not joined by an edge, then there is a set of vertices whose removal leaves a graph in which *s* and *t* lie in different pieces. What is the smallest number of vertices in such a set?

On the other hand, there may be several paths joining *s* and *t* that share no vertices other than *s* and *t*. How many such paths can there be? Clearly this number cannot exceed the number of vertices whose removal leaves *s* and *t* in different pieces. That this many paths always exist is the essence of the fundamental theorem of connectivity, first given by Menger [33] in 1927:

For any non-adjacent vertices s and t in a graph, the maximum number of internally disjoint paths from s to t equals the minimum number of vertices in a set whose removal leaves s and t in different pieces.

For example, in the graph below there are three vertices whose removal separates *s* from *t*, and there are three paths from *s* to *t* that are internally disjoint.



There are numerous versions of Menger's theorem, some of which are global in nature. A graph is *k-connected* if the removal of any set of fewer than *k* vertices leaves a connected graph with at least two vertices. Here is a global form of Menger's result:

If a graph is k-connected, then each pair of vertices has a family of k internally disjoint paths between them.

The *connectivity* of a graph is the minimum number of vertices whose removal results in a disconnected graph (or leaves just one vertex, in the case of a complete graph). Although Menger's theorem does not tell us how to find the connectivity of a graph efficiently, there are methods for doing so, and we return to this point shortly.

The connectivity of a graph has practical implications, such as determining how vulnerable to disruption a communications network might be if certain sites break down. Alternatively, instead of sites it could be links that fail, and this suggests edge versions of connectedness. Unsurprisingly, there are both local and global forms of Menger's theorem for edges; first a local version:

For any vertices s and t in a graph, the maximum number of edge-disjoint paths from s to t equals the minimum number of edges in a set whose removal leaves s and t in different pieces.

A graph is *k-edge-connected* if the removal of any set of fewer than *k* edges leaves a connected graph with at least two vertices. Here is a global form of Menger's result for edges:

If a graph is k-edge-connected, then each pair of vertices has a family of k edge-disjoint paths between them.

The *edge-connectivity* of a graph is defined as expected, and is a measure of how vulnerable a network might be to disruption if lines of communication fail.

Finally, we remark that Ford and Fulkerson's max-flow–min-cut theorem is actually equivalent to Menger's theorem, and thus implies all of its variations. This is another illustration of something that we have seen before (for example, in connection with matchings and factorizations), that there have been independent discoveries, often many years later and through different applications, of closely related ideas and results. Furthermore, the proof that Ford and Fulkerson provided is constructive in nature, and thus provides an efficient method for finding both a flow of maximum value and a cut of minimum capacity in a network. This notion of an efficient constructive method leads us to our final stop in this excursion into the history of modern graph theory.

Algorithmic graph theory

Graph theory algorithms can be traced back over one hundred years to when, for example, M. Trémaux and others explained how to escape from a maze (see [31]). In the mid 20th century such algorithms increasingly came into their own, with the solutions of such problems as the *shortest path problem*, the *minimum connector problem*, and the *Chinese postman problem*. In each of these problems we are given a network or weighted graph with a number assigned to each edge, such as its length or the time taken to traverse it.

There are several efficient algorithms for finding the shortest path in a given network, of which the best known is due to E. W. Dijkstra [13] in 1959. Finding a longest path, or *critical path*, in an activity network also dates from the 1940s and 1950s, with PERT (Program Evaluation and Review Technique) used by the US Navy for problems involving the building of submarines, and CPM (Critical Path Method) developed by the Du Pont de Nemours Company to minimize the total cost of a project. The *Chinese postman problem* is to find the shortest route that covers each edge of a given weighted graph; it was solved by Meigu Guan [Mei-Ku Kwan] [19]. The greedy algorithm for the *minimum connector problem*, in which we seek a minimum-length spanning tree in a weighted graph, can be traced back to O. Borůvka [9] and was later rediscovered by J. B. Kruskal [28].

A problem that sounds similar is the *travelling salesman problem*, in which a salesman wishes to make a tour of a number of cities in minimum time or distance (see [30]). This problem appeared in rudimentary form in a practical book written in 1831 for the *Handlungsreisende* (commercial traveller) (see

Voigt [48]), but its first appearance in mathematical circles was not until the early 1930s, at Princeton. It was later popularized at the RAND Corporation (see [11]), eventually leading to a fundamental paper of Dantzig *et al.* [12] that included the solution of a travelling salesman problem with forty-nine cities. Over the years the number of cities has gradually increased, and in the 1980s a problem with more than two thousand cities was settled by Padberg and Rinaldi [34]. It can now be solved for tens of thousands of cities (see [4]).

The travelling salesman problem was not the only significant combinatorial problem studied at the RAND Corporation in the mid 20th century. In particular, algorithms were developed by Dantzig and by Ford and Fulkerson for finding the maximum flow of a commodity between two nodes in a capacitated network (see the previous section). Algorithms for solving matching and assignment problems were also developed, where one wishes to assign people as appropriately as possible to jobs for which they are qualified; this work developed from the above-mentioned work of König and from the celebrated result on matching due to Philip Hall [20], later known as the 'marriage theorem' (see Chapter 13).

This chapter would not be complete without some mention of one of the outstanding problems of 20th-century mathematics:

Is there a 'good' algorithm for solving the travelling salesman problem?

A class of problems is said to be in P if any instance of it can be solved in a number of steps bounded by a polynomial in the input size (for graphs, usually the number of vertices); it is in NP if a proposed solution can be checked in a polynomial number of steps. The basic question is 'Does P = NP?'

In a groundbreaking paper in 1975, R. Karp [26] presented a number of important graph theory problems with the property that if any one of them has a 'good' algorithm for its solution, then so do all of the others; these problems include determining whether a graph is 3-colourable, whether two drawings represent the same graph, whether a graph has a Hamiltonian cycle (see Chapter 8), and many hundreds of others. The 'P = NP?' question remains unresolved and was one of the Millennium Problems posed by the Clay Mathematics Institute in 2000, with a prize of one million US dollars for a solution.

References

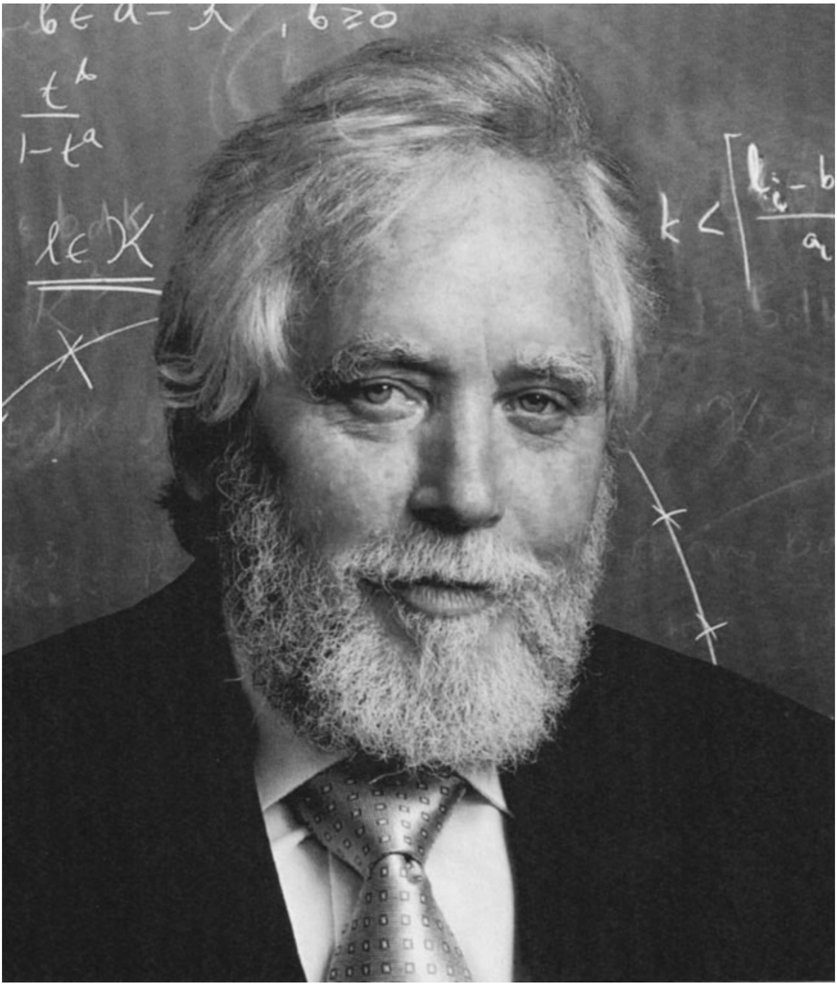
1. An earlier version of part of this chapter appears in N. L. Biggs, E. K. Lloyd, and R. J. Wilson, The history of combinatorics, *Handbook of Combinatorics* (ed. R. Graham *et al.*), Elsevier (1995), Chapter 44.
2. K. Appel and W. Haken, The solution of the four-color-map problem, *Sci. Amer.* 237, No. 4 (October 1977), 108–21.
3. K. Appel and W. Haken, Every planar map is four colorable: Part 1, Discharging, Part 2 (with J. Koch) Reducibility, *Illinois J. Math*. 21 (1977), 429–90 and 491–567.
4. D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*, Princeton University Press (2007).
5. N. L. Biggs, E. K. Lloyd, and R. J. Wilson, *Graph Theory 1736–1936*, Clarendon Press, Oxford (1976, 1998).
6. G. D. Birkhoff, A determinantal formula for the number of ways of coloring a map, *Ann. Math*. 14 (1912), 42–6.
7. G. D. Birkhoff, The reducibility of maps, *Amer. J. Math*. 35 (1913), 115–28.
8. G. D. Birkhoff and D. C. Lewis, Chromatic polynomials, *Trans. Amer. Math. Soc*. 60 (1946), 355–451.
9. O. Borůvka, O jistém problému minimálním, *Acta Soc. Sci. Natur. Moravicae* 3 (1926), 37–58.
10. R. L. Brooks, On colouring the nodes of a network, *Proc. Cambridge Philos. Soc*. 37 (1941), 194–7.
11. G. B. Dantzig, Reminiscences about the origins of linear programming, *Oper. Res. Lett*. 1 (1982), 43–8.
12. G. B. Dantzig, D. R. Fulkerson, and S. M. Johnson, Solution of a large-scale traveling-salesman problem, *Oper. Res*. 2 (1954), 393–410.
13. E. W. Dijkstra, A note on two problems in connexion with graphs, *Numer. Math*. 1 (1959), 269–71.
14. H. E. Dudeney, Perplexities, *Strand Mag*. 46 (July 1913), 110 and 46 (August 1913), 221.
15. L. R. Ford and D. R. Fulkerson, Maximal flow through a network, *Canad. J. Math*. 8 (1956), 399–404.
16. P. Franklin, The four color problem, *Amer. J. Math*. 44 (1922), 225–36.
17. P. Franklin, A six color problem, *J. Math. Phys*. 13 (1934), 363–9.
18. H. H. Glover, J. P. Huneke, and C. S. Wang, 103 graphs that are irreducible for the projective plane, *J. Combin. Theory (B)* 27 (1979), 332–70.
19. Meigu Guan (= Mei-Ku Kwan), Graphic programming using odd or even points, *Acta Math. Sinica* 10 (1960), 263–6 = *Chinese Math*. 1 (1962), 273–7.
20. P. Hall, On representatives of subsets, *J. London Math. Soc*. 10 (1935), 26–30.
21. P. J. Heawood, Map-colour theorem, *Quart. J. Pure Appl. Math*. 24 (1890), 332–8.
22. H. Heesch, *Untersuchungen zum Vierfarbenproblem*, B. I. Hochschulscripten, 810/810a/810b (1969).

23. L. Heffter, Über das Problem der Nachbargebiete, *Math. Ann*. 38 (1891), 477–508.
24. T. R. Jensen and B. Toft, *Graph Coloring Problems*, Wiley-Interscience (1995).
25. I. N. Kagno, A note on the Heawood color formula, *J. Math. Phys*. 14 (1935), 228–31.
26. R. Karp, On the computational complexity of combinatorial problems, *Networks* 5 (1975), 45–68.
27. D. König, Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, *Math. Ann*. 77 (1916), 453–65.
28. J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem, *Proc. Amer. Math. Soc*. 7 (1956), 48–50.
29. C. Kuratowski, Sur le problème des courbes gauches en topologie, *Fund. Math*. 15 (1930), 271–83.
30. E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Schmoys (eds), *The Traveling Salesman Problem: A Guided Tour through Combinatorial Optimization*, Wiley (1985).
31. E. Lucas, *Récréations Mathématiques*, Vol. I, Gauthier-Villars (1882), 47–50 (reprinted by Albert Blanchard, 1973).
32. J. Mayer, Le problème des régions voisines sur les surfaces closes orientables, *J. Combin. Theory* 6 (1969), 177–95.
33. K. Menger, Zur allgemeinen Kurventheorie, *Fund. Math*. 10 (1927), 96–115.
34. M. W. Padberg and G. Rinaldi, Optimization of a 532-city symmetric traveling salesman problem by branch and cut, *Oper. Res. Lett*. 6 (1987), 1–7.
35. J. Petersen, Die Theorie der regulären Graphs, *Acta Math*. 15 (1891), 193–220.
36. J. Petersen, [Sur le théorème de Tait], *Interméd. Math*. 5 (1898), 225–7.
37. G. Ringel, *Map Color Theorem*, Springer-Verlag (1974).
38. N. Robertson and P. D. Seymour, Graph minors – a survey, *Surveys in Combinatorics 1985* (ed. I. Anderson), London Mathematical Society Lecture Notes 103, Cambridge University Press (1985), 153–71.
39. N. Robertson, D. Sanders, P. Seymour, and R. Thomas, The four-colour theorem, *J. Combin. Theory (B)* 70 (1997), 2–44.
40. C. E. Shannon, A theorem on coloring the lines of a network, *J. Math. Phys*. 28 (1949), 148–51.
41. P. G. Tait, On the colouring of maps, *Proc. Roy. Soc. Edinburgh* 10 (1878–80), 501–3.
42. R. Thomas, An update on the four-colour theorem, *Notices Amer. Math. Soc*. 45(7) (August 1998), 848–59.
43. H. Tietze, Einige Bemerkungen über das Problem des Kartenfärbens auf einseitigen Flächen, *Jber. Deut. Math.-Ver*. 19 (1910), 155–79.
44. W. T. Tutte, The factorization of linear graphs, *J. London Math. Soc*. 22 (1947), 107–11.
45. W. T. Tutte, The factors of graphs, *Canad. J. Math*. 4 (1952), 314–28.
46. W. T. Tutte, Matroids and graphs, *Trans. Amer. Math. Soc*. 90 (1959), 527–52.
47. V. G. Vizing, On an estimate of the chromatic class of a *p*-graph, *Diskret. Analiz* 3 (1964), 25–30.

48. B. F. Voigt, *Der Handlungsreisende, wie er sein soll und was er zu thun hat, um Aufträge zu erhalten und einer glücklichen Erfolgs in seinen Geschäften gewiss zu sein* (*von einem alten Commis-Voyageur, Ilmenau*) (1831) (reprinted by Schramm, 1981).
49. P. Wernicke, Über den kartographischen Vierfarbensatz, *Math. Ann.* 58 (1904), 413–26.
50. H. Whitney, The coloring of graphs, *Proc. Nat. Acad. Sci. U. S. A.* 17 (1931), 122–5.
51. H. Whitney, Non-separable and planar graphs, *Proc. Nat. Acad. Sci. U. S. A.* 17 (1931), 125–7.
52. H. Whitney, On the abstract properties of linear dependence, *Amer. J. Math.* 57 (1935), 509–33.
53. R. Wilson, *Four Colours Suffice*, Allan Lane (2002) (revised color edn., Princeton University Press, 2014).

PART IV

AFTERMATH



Endre Szemerédi (b. 1940), winner of the Abel Prize, 2012.

A PERSONAL VIEW OF COMBINATORICS

PETER J. CAMERON

This chapter presents a quick overview of the recent development of combinatorics and its current directions – as a discipline in its own right, as a part of mathematics and, more generally, as a part of science and of society.

Introduction

Henry Whitehead reportedly said, 'Combinatorics is the slums of topology' (an attribution confirmed by Graham Higman, a student of Whitehead). Less disparagingly, Hollingdale [24] wrote '. . . the branch of topology we now call "graph theory" . . . '. This prejudice, the view that combinatorics is quite different from 'real mathematics', was not uncommon in the 20th century, among popular expositors as well as professionals. In his biography of Srinivasa Ramanujan, Robert Kanigel [26] describes Percy MacMahon in these terms:

> [MacMahon's] expertise lay in combinatorics, a sort of glorified dice-throwing, and in it he had made contributions original enough to be named a Fellow of the Royal Society.

In the later part of the century, attitudes changed. When the 1998 film *Good Will Hunting* featured a famous mathematician at the Massachusetts Institute of Technology who had won a Fields Medal for combinatorics, many found this somewhat unbelievable; nonetheless, the 'unsolvable math problem' in this film is based on the actual experience of George B. Dantzig, who as a student solved two problems posed by Jerzy Neyman at Berkeley in 1940 (see [8]). In this case, however, life followed art later in 1998 when Fields Medals were awarded to Tim Gowers and Richard Borcherds for their work, much of which was in combinatorics (though John McKay points out that Borcherds would probably not regard his work as being combinatorial). In 2012, the Abel prizewinner was Endre Szemerédi, the distinguished Hungarian combinatorialist.

A more remarkable instance of life following art involves Stanisław Lem's 1968 novel *His Master's Voice* [33]. The narrator, a mathematician, describes how he single-mindedly attacked his rival's work:

> I do not think I ever finished any larger paper in all my younger work without imagining Dill's eyes on the manuscript. What effort it cost me to prove that the Dill variable combinatorics was only a rough approximation of an ergodic theorem! Not before or since, I daresay, did I polish a thing so carefully; and it is even possible that the whole concept of groups later called Hogarth groups came out of that quiet, constant passion with which I plowed Dill's axioms under.

In 1975, Szemerédi [46] published his remarkable combinatorial proof that a set of natural numbers with positive density contains arbitrarily long arithmetic progressions; in 1977, Furstenberg [15] gave a proof based on ergodic theory! (This is not to suggest that Furstenberg's attitude to Szemerédi parallels Hogarth's to Dill in the novel.)

In this chapter, I have attempted to tease apart some of the interrelated reasons for this change, and perhaps to throw some light on present trends and future directions. I have divided the causes into four groups:

- the influence of the computer;
- the growing sophistication of combinatorics;
- its strengthening links with the rest of mathematics;
- wider changes in society.

I have told the story mostly through examples.

The influence of the computer

Even before computers were built, pioneers such as Charles Babbage and Alan Turing realized that they would be designed on discrete principles, and would raise theoretical issues that would lead to important mathematics.

Kurt Gödel [18] showed that there are true statements about the natural numbers that cannot be deduced from the axioms of a standard system such as Peano's. This result was highly significant for the foundations of mathematics, but Gödel's unprovable statement itself had no significance in the actual practice of mathematics. The first example of a natural mathematical statement which is unprovable in Peano arithmetic was discovered by Paris and Harrington [36], and is a theorem in combinatorics (it is a slight strengthening of Ramsey's theorem). It is unprovable from the axioms because the corresponding 'Paris–Harrington function' grows faster than any provably computable function. Several further examples of this phenomenon have been discovered, mostly combinatorial in nature. For instance, calculating precise values for Ramsey numbers, or even close estimates, appears to be among the most fiendishly difficult open combinatorial problems.

More recently, attention has turned from *computability* to *computational complexity*: given that something can be computed, what resources (time, memory, etc.) are required for the computation? Recall (from Chapter 14) that a class of problems is said to be *polynomial-time computable*, or in P, if any instance can be solved in a number of steps bounded by a polynomial in the input size. A class is in NP if the same assertion holds if we are allowed to make a number of lucky guesses (or, what amounts to the same thing, if a proposed solution can be checked in a polynomial number of steps). The great unsolved problem of complexity theory asks:

$$\text{Is P} = \text{NP?}$$

On 24 May 2000, the Clay Mathematics Institute announced a list of seven unsolved problems, for each of which a prize of one million dollars was offered. The P = NP? problem was the first on the list [9]. This problem is particularly important for combinatorics, since many hundreds of intractable combinatorial problems (including the existence of a Hamiltonian cycle in a graph) are known to be in NP. In the unlikely event of an affirmative answer, 'fast' algorithms would exist for all these problems.

Now we turn to the practical use of computers. Computer systems, such as GAP [16], have also been developed which can treat algebraic or combinatorial objects, such as a group or a graph, in a way similar to the handling of complex numbers or matrices in more traditional systems. These give the mathematician a very powerful tool for exploring structures and testing (or even formulating) conjectures.

But what has caught the public eye is the use of computers to prove theorems. This was dramatically the case in 1976 when Kenneth Appel and Wolfgang Haken [1] announced that they had proved the four-colour theorem by computer (see Chapter 14). Their announcement started a wide discussion over whether a computer proof is really a 'proof' at all: see, for example, Swart [45] and Tymoczko [50] for contemporary responses, and Wilson [52] for a survey. An even more massive computation by Clement Lam and his co-workers [29], discussed by Lam in [28], showed the non-existence of a projective plane of order 10 (see Chapters 10 and 11). Other recent achievements include the classification of Steiner triple systems of order 19 (see [27]).

Computers have also been used in other parts of mathematics. For example, in the classification of finite simple groups (discussed below), many of the sporadic simple groups were constructed with the help of computers. The very practical study of fluid dynamics depends on massive computation. What distinguishes combinatorics? Two factors seem important:

- in a sense, the effort of the proof consists mainly in detailed case analysis, or generates large amounts of data, and so the computer does most of the work;
- the problem and solution are both discrete; the results are not invalidated by rounding errors or chaotic behaviour.

Finally, the advent of computers has given rise to many new areas of mathematics related to the processing and transmission of data. Since computers are digital, these areas are naturally related to combinatorics. They include coding theory (discussed below), cryptography, integer programming, discrete optimization, and constraint satisfaction.

The nature of the subject

The last two centuries of mathematics have been dominated by the trend towards axiomatization. A structure which fails to satisfy the axioms is not to be considered. (As one of my colleagues put it to a student in a class, 'For a ring to pass the exam, it has to get 100%.') Combinatorics has never fitted this pattern very well.

When Gian-Carlo Rota and various co-workers wrote an influential series of papers with the title 'On the foundations of combinatorial theory' in the 1960s and 1970s (see [40] and [10], for example), one reviewer compared combinatorialists to nomads on the steppes who had not managed to construct the cities in which other mathematicians dwell, and expressed the hope that these papers would at least found a thriving settlement.

While Rota's papers have been very influential, this view has not prevailed. To see this, we turn to the more recent series on 'Graph minors' by Neil Robertson and Paul Seymour [39]. These are devoted to the proof of a single major theorem, that a minor-closed class of graphs is determined by finitely many excluded minors. Along the way, a rich tapestry is woven, which is descriptive (giving a topological embedding of graphs) and algorithmic (showing that many graph problems lie in P), as well as deductive.

The work of Robertson and Seymour and its continuation is certainly one of the major themes in graph theory at present, and has contributed to a shorter proof of the four-colour theorem, as well as to a proof of the strong perfect graph conjecture. Various authors – notably Gerards, Geelen, and Whittle – are extending it to classes of matroids (see [17]).

What is clear, though, is that combinatorics will continue to elude attempts at formal specification.

Relationships

In 1974 an Advanced Study Institute on Combinatorics was held at Nijenrode in the Netherlands, organized by Marshall Hall and Jack van Lint. This was one of the first presentations, aimed at young researchers, of combinatorics as a mature mathematical discipline. The subject was divided into five sections: the theory of designs, finite geometry, coding theory, graph theory, and combinatorial group theory.

It is very striking to look at the four papers in coding theory [22]. This was the youngest of the sections, having begun with the work of Hamming and Golay in the late 1940s. Yet the methods being used involved the most sophisticated mathematics: invariant theory, harmonic analysis, Gauss sums, and Diophantine equations.

This trend has continued. In the 1970s the Russian school (notably Goppa, Manin, and Vladut) developed links between coding theory and algebraic geometry (specifically, divisors on algebraic curves); these links were definitely 'two-way', and both subjects benefited. More recently, codes over rings and quantum codes have revitalized the subject and made new connections with ring theory and group theory. In the related field of cryptography, one of the most widely used ciphers is based on elliptic curves.

Another example is provided by the most exciting development in mathematics in the late 1980s, which grew from the work of Vaughan Jones, for which he received a Fields Medal in 1990. His research on traces of Von Neumann algebras came together with representations of the Artin braid group to yield a new invariant of knots, with ramifications in mathematical physics and elsewhere (see the citation by Joan Birman [3] and her popular account [4] for a map of this territory).

Later, it was pointed out that the Jones polynomial is a specialization of the Tutte polynomial, which had been defined for arbitrary graphs by Tutte and Whitney and generalized to matroids by Tutte; Tutte himself gave two accounts of his discovery (see [48] and [49]). The connections led to further researches. There was the work of François Jaeger [25], who derived a spin model, and hence an evaluation of the Kauffman polynomial, from the strongly regular graph associated with the Higman–Sims simple group; and also the work of Dominic Welsh and his collaborators (described in his book [51]) on the computational complexity of the new knot invariants.

Sokal's article [44] discusses the close relations between the Tutte polynomial and the partition function for the Potts model in statistical mechanics. This interaction has led to important advances in both areas. The connection actually goes back to Fortuin and Kasteleyn [14], with later contributions by Zaslavsky and by Bollobás and Riordan. (I am grateful to Jo Ellis-Monaghan for information about this.)

By their nature, examples such as this of unexpected connections cannot be predicted. However, combinatorics is likely to be involved in such discoveries: it seems that deep links in mathematics often reveal themselves in combinatorial patterns.

One of the best examples concerns the ubiquity of the Coxeter–Dynkin diagrams A_n, D_n, E_6, E_7, and E_8. To guide the development of mathematics, Arnol'd (see [7]) proposed finding an explanation of their ubiquity as a modern equivalent of a Hilbert problem. He noted their occurrence in areas such as Lie algebras (the simple Lie algebras over the complex numbers), Euclidean geometry (root systems), group theory (Coxeter groups), representation theory (algebras of finite representation type), and singularity theory (singularities with definite intersection form), as well as their connection with the regular polyhedra. To this list could be added mathematical physics (instantons) and combinatorics (eigenvalues of graphs). Indeed, graph theory provides the most striking specification of the diagrams: they are just the connected graphs with all eigenvalues smaller than 2.

Recently this subject has been revived with the discovery by Fomin and Zelevinsky [13] of the role of these ADE diagrams in the theory of cluster algebras: this is a new topic with combinatorial foundations and applications in Poisson geometry, integrable systems, representation theory, and total positivity.

Other developments include the relationship of combinatorics to finite group theory. The classification of finite simple groups [19] is the greatest collaborative effort ever in mathematics, running to about 15 000 journal pages. (Ironically, although the theorem was announced in 1980, the proof contained a gap which has only just been filled.) Combinatorial ideas (graphs, designs, codes, and geometries) were involved in the proof – perhaps, most notably, the classification of spherical buildings by Jacques Tits [47]. Also, the result has had a great impact in combinatorics, with consequences both for symmetric objects such as graphs and designs (see Praeger's survey [38]), and (more surprisingly) elsewhere, as in Luks's proof [34] that the graph isomorphism problem for graphs of bounded valency is in P.

This account would not be complete without a mention of the work of Borcherds [5] on 'monstrous moonshine', connecting the Golay code, the Leech lattice, and the Monster simple group with generalized Kac–Moody algebras and vertex operators in mathematical physics, and throwing up a number of product identities of the kind familiar from the classical work of Jacobi and others.

In science and in society

Like any human endeavour, combinatorics has been affected by the great changes in society during the last century. The first influence to be mentioned is a single individual, Paul Erdős, who is the subject of two recent best-selling biographies [23], [41].

Erdős's mathematical interests were wide, but combinatorics was central to them. As we saw in Chapter 13, he spent a large part of his life without a permanent abode, travelling the world and collaborating with hundreds of mathematicians. In the days before electronic mail he was a vital communication link between mathematicians in the East and the West. He also inspired a vast body of research – his more than 1500 papers dwarf the output of any other modern mathematician (see [21]).

Erdős also stimulated mathematics by publicizing his vast collection of problems; for many of them, he offered financial rewards for solutions. As an example, here is one of his most valuable problems:

> Let $A = \{a_1, a_2, \ldots\}$ be a set of positive integers with the property that the sum of the reciprocals of the members of A diverges. Is it true that A contains arbitrarily long arithmetic progressions?

The motivating special case (recently solved in the affirmative by Green and Tao [20]) is where A is the set of prime numbers; this is a problem in number theory, but Erdős's extension to an arbitrary set transforms it into combinatorics.

Increased collaboration among mathematicians goes beyond the influence of Erdős; combinatorics seems to lead the trend. Aspects of this trend include large international conferences (the Southeastern Conference on Combinatorics, Graph Theory, and Computing, which held its 43rd meeting in 2012, attracts over five hundred people annually), and electronic journals (the *Electronic Journal of Combinatorics* [12], founded in 1994, was one of the first refereed specialist electronic journals in mathematics). Electronic publishing is particularly attractive to combinatorialists; often, arguments require long case analysis, which editors of traditional print journals may be reluctant to include in full.

On a popular level, the sudoku puzzle (a variant of the problem of completing a critical set in a Latin square) daily engages many thousands of people throughout the world in combinatorial reasoning (see Chapter 11). Mathematicians have not been immune to its attractions. At the time of writing, MathSciNet lists over forty publications with 'sudoku' in their title, linking it to topics as diverse as spreads and reguli, neural networks, fractals, and Shannon entropy.

Our time has seen a change in the scientific viewpoint from the continuous to the discrete. Two mathematical developments of the 20th century – catastrophe theory and chaos theory – have shown how discrete effects can be produced by continuous causes. (Perhaps their dramatic names reflect the intellectual shock of this discovery.) But the trend has been even more widespread. In their 1944 book introducing a new branch of discrete mathematics (game theory), John von Neumann and Oskar Morgenstern [35] wrote:

> The emphasis on mathematical methods seems to be shifted more towards combinatorics and set theory – and away from the algorithm of differential equations which dominates mathematical physics.

How does discreteness arise in nature? Segerstråle [42] quotes John Maynard Smith as saying:

> today we really do have a mathematics for thinking about complex systems and things which undergo transformations from quantity into quality

or from continuous to discrete, mentioning Hopf bifurcations as a mechanism for this.

On the importance of discreteness in nature, Steven Pinker [37] has no doubt. He wrote:

> It may not be a coincidence that the two systems in the universe that most impress us with their open-ended complex design – life and mind – are based on discrete combinatorial systems.

Here, 'mind' refers primarily to language, whose combinatorial structure is well described in Pinker's book; 'life' refers to the genetic code, where DNA molecules can be regarded as words in an alphabet of four letters (the bases adenine, cytosine, guanine, and thymine), and three-letter subwords encode amino acids, the building blocks of proteins.

The Human Genome Project, whose completion was announced in 2001, was a major scientific enterprise designed to describe completely the genetic code of humans (see [2] for an account of the mathematics involved, and [32] for subsequent developments). At Pinker's institution (the Massachusetts Institute of Technology), the Whitehead Laboratory was engaged in this project. Its director, Eric Lander, serves to round off this final chapter and illustrate its major themes. His doctoral thesis [30] was in combinatorics, involving a 'modern' subject (coding theory), links within combinatorics (codes and designs), and links to other parts of mathematics (lattices and local fields), and furthermore he is a fourth-generation academic descendant of Henry Whitehead. Eric Lander honoured me at my 60th birthday conference by giving a talk entitled 'The human genome: an asymmetric design'; the title was a parody of the book he wrote based on his thesis: *Symmetric Designs: An Algebraic Approach* [31].

But there are now hints that discreteness plays an even more fundamental role. One of the goals of physics at present is the construction of a theory which could reconcile the two pillars of 20th-century physics: general relativity and quantum mechanics. In describing string theory, loop quantum gravity, and a variety of other approaches (including non-commutative geometry and causal set theory), Smolin [43] argued that all of them involve discreteness at a fundamental level (roughly, the Planck scale, which is much too small and fleeting to be directly observed). Causal set theory is based on discrete partially ordered sets and has already attracted the attention of combinatorialists (see [6] and [11]). Indeed, developments such as the holographic principle suggest that the basic currency of the universe may not be space and time, but information measured in bits. Maybe the 'theory of everything' will be combinatorial!

References

1. K. Appel and W. Haken, Every planar map is four colorable, *Bull. Amer. Math. Soc.* 82 (1976), 711–12.
2. S. Batzoglou, B. Berger, D. J. Kleitman, E. S. Lander, and L. Pachter, Recent developments in computational gene recognition, *Proc. International Congress of Mathematicians,* Berlin, 1998*, Documenta Mathematica*, Extra Vol. I, 649–58 (electronic).
3. J. S. Birman, The work of Vaughan F. R. Jones, *Proc. International Congress of Mathematicians*, Kyoto, 1990, Mathematical Society of Japan/Springer-Verlag (1991), 9–18.
4. J. S. Birman, Recent developments in braid and link theory, *Math. Intelligencer* 13 (1991), 52–60.
5. R. Borcherds, What is moonshine?, *Proc. International Congress of Mathematicians*, Berlin, 1998, *Documenta Mathematica*, Extra Vol. I, 607–15 (electronic).
6. G. Brightwell and M. Luczak, Order-invariant measures on causal sets, *Ann. Appl. Probab.* 21 (2011), 1493–1536.
7. F. E. Browder, Problems of present day mathematics, *Mathematical Developments Arising from Hilbert Problems*, Proc. Symposia in Pure Mathematics 28, American Mathematical Society (1976), 35–79.
8. J. H. Brunvald, *Too Good to be True: The Colossal Book of Urban Legends*, W. W. Norton (1999).
9. Clay Mathematics Institute, P vs NP problem, http://www.claymath.org/millennium/P_vs_NP/
10. H. Crapo and G.-C. Rota, *On the Foundations of Combinatorial Theory: Combinatorial Geometries*, MIT Press (1970).
11. M. Droste, Universal homogeneous causal sets, *J. Math. Phys.* 46 (2005), no. 12, 122503, 10 pp.
12. *Electronic Journal of Combinatorics*, http://www.combinatorics.org/
13. S. Fomin and A. Zelevinsky, Cluster algebras, II: Finite type classification, *Inventiones Math.* 154 (2003), 53–121.
14. C. M. Fortuin and P. W. Kasteleyn, On the random-cluster model, I: Introduction and relation to other models, *Physica* 57 (1972), 536–64.
15. H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* 31 (1977), 204–56.
16. GAP homepage, http://www.gap-system.org/

17. B. Gerards, J. Geelen, and G. Whittle, Towards a matroid-minor structure theory, *Combinatorics, Complexity, and Chance* (ed. G. Grimmett and C. McDiarmid), Oxford Lecture Series in Mathematics and Its Applications 34, Oxford University Press (2007), 72–82.
18. K. Gödel, Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, *Monatsh. Math. Phys.* 38 (1931), 173–98.
19. D. Gorenstein, *Finite Simple Groups: An Introduction to Their Classification*, Plenum Press (1982).
20. B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Ann. Math.* (2) 167 (2008), 481–547.
21. J. W. Grossman, Paul Erdős: the master of collaboration, *The Mathematics of Paul Erdős* (ed. R. L. Graham and J. Nešetřil), Springer-Verlag (1997), 467–75.
22. M. Hall, Jr., and J. H. van Lint, *Combinatorics*, NATO ASI Series C16, D. Reidel (1975).
23. P. Hoffman, *The Man Who Loved Only Numbers*, Fourth Estate (1998).
24. S. Hollingdale, *Makers of Mathematics*, Penguin (1989).
25. F. Jaeger, Strongly regular graphs and spin models for the Kauffman polynomial, *Geometriae Dedicata* 44 (1992), 23–52.
26. R. Kanigel, *The Man Who Knew Infinity: A Life of the Genius Ramanujan*, Charles Scribner's Sons (1991).
27. P. Kaski and P. Østergård, The Steiner triple systems of order 19, *Math. Comput.* 73 (2004), 2075–92.
28. C. W. H. Lam, The search for a finite projective plane of order 10, *Amer. Math. Monthly* 98 (1991), 305–18.
29. C. W. H. Lam, S. Swiercz, and L. Thiel, The nonexistence of finite projective planes of order 10, *Canad. J. Math.* 41 (1989), 1117–23.
30. E. S. Lander, *Topics in Algebraic Coding Theory*, DPhil thesis, University of Oxford (1980).
31. E. S. Lander, *Symmetric Designs: An Algebraic Approach*, London Mathematical Society Lecture Notes 74, Cambridge University Press (1983).
32. E. S. Lander, Initial impact of the sequencing of the human genome, *Nature* 470 (2011), 187–97.
33. S. Lem, *His Master's Voice*, Harvest Books (1984).
34. E. M. Luks, Isomorphism of graphs of bounded valence can be tested in polynomial time, *J. Comput. Syst. Sci.* 25 (1982), 42–65.
35. J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press (1944).
36. J. Paris and L. Harrington, A natural incompleteness in Peano arithmetic, *Handbook of Mathematical Logic* (ed. J. Barwise), North-Holland (1977), 1133–42.
37. S. Pinker, *The Language Instinct*, Penguin (1994).
38. C. E. Praeger, Finite transitive permutation groups and finite vertex-transitive graphs, *Graph Symmetry* (ed. G. Hahn and G. Sabidussi), Kluwer (1997), 277–318.
39. N. Robertson and P. D. Seymour, Graph minors, I: Excluding a forest, *J. Combin. Theory (B)* 35 (1983), 39–61, and subsequent papers with the same title.

40. G.-C. Rota, On the foundations of combinatorial theory, I: Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie* 2 (1964), 340–68.
41. B. Schechter, *My Brain is Open: The Mathematical Journeys of Paul Erdős*, Simon & Schuster (1998).
42. U. Segerstråle, *Defenders of the Truth: The Battle for Science in the Sociobiology Debate and Beyond*, Oxford University Press (2000).
43. L. Smolin, *Three Roads to Quantum Gravity*, Weidenfeld & Nicholson (2000).
44. A. D. Sokal, The multivariate Tutte polynomial (alias Potts model) for graphs and matroids, *Surveys in Combinatorics 2005* (ed. B. S. Webb), London Mathematical Society Lecture Notes 327, Cambridge University Press (2005), 173–226.
45. E. R. Swart, The philosophical implications of the four-color problem, *Amer. Math. Monthly* 87 (1980), 697–707.
46. E. Szemerédi, On sets of integers containing no *k* elements in arithmetic progression, *Proc. International Congress of Mathematicians*, Vancouver, 1974, Vol. 2, Canadian Math. Congress, Montreal (1975), 503–5.
47. J. Tits, *Buildings of Spherical Type and Finite BN-Pairs*, Lecture Notes in Mathematics 386, Springer (1974).
48. W. T. Tutte, *Graph Theory as I Have Known It*, Oxford University Press (1998).
49. W. T. Tutte, The coming of the matroids, *Surveys in Combinatorics, 1999* (ed. J. D. Lamb and D. A. Preece), London Mathematical Society Lecture Notes 267, Cambridge University Press (1999), 3–20.
50. T. Tymoczko, Computers, proofs and mathematicians: A philosophical investigation of the four-color proof, *Math. Mag.* 53 (1980), 131–8.
51. D. J. A. Welsh, *Complexity: Knots, Colourings and Counting*, London Mathematical Society Lecture Notes 186, Cambridge University Press (1993).
52. R. Wilson, *Four Colours Suffice*, Allen Lane (2002) (revised color edn., Princeton University Press, 2014).

NOTES ON CONTRIBUTORS

Lars Døvling Andersen is Professor of Mathematics at Aalborg University, where he has been since 1986. He received his PhD degree from the University of Reading, and his main combinatorial interests have been graph theory and design theory. Several of his papers involve the interplay between these two areas, using graph-theoretic techniques such as edge-colouring and amalgamation in design theory. He was a managing editor of the *Journal of Graph Theory* and has served on the European Mathematical Society's Committee for Developing Countries. He has held a Niels Bohr Fellowship and was awarded the Hans Gram medal by the Royal Danish Academy of Sciences and Letters.

Ian Anderson is an Honorary Research Fellow at the University of Glasgow, where he taught mathematics for 40 years, introducing several undergraduate combinatorial courses. He is the author of a number of books on combinatorial mathematics, including *Combinatorics of Finite Sets* and *Combinatorial Designs and Tournaments*. He has also written several papers dealing with the early history of combinatorial designs, and he co-authored the historical survey 'Design theory: antiquity to 1950' in the second edition of the *CRC Handbook of Combinatorial Designs*. He served as a member of the British Combinatorial Committee for 14 years.

George E. Andrews is Evan Pugh Professor of Mathematics at Pennsylvania State University and is an expert on the theory of partitions. He has a long-term interest in the work of S. Ramanujan, whose 'lost notebook' he unearthed in 1976, and is now collaborating with Bruce Berndt on a series of volumes explicating the brilliant and sometimes enigmatic ideas in this notebook. He was elected to the American Academy of Arts and Sciences in 1997, and to the National Academy of Sciences (USA) in 2003. In 2009 he became a Fellow of the Society of Industrial and Applied Mathematics, and in 2012 a Fellow of the American Mathematical Society. He holds honorary degrees from the Universities of Parma, Florida, and Waterloo. He has received many awards for his teaching and service to the profession, and was President of the American Mathematical Society from 2009 to 2011.

Lowell Beineke received a Bachelor of Science degree in mathematics from Purdue University before earning his doctorate in mathematics from the University of Michigan with a dissertation on the thickness of graphs. He has spent his academic career at Indiana University–Purdue University Fort Wayne, receiving honours for both research and teaching. Awarded the Schrey Professorship of Mathematics in 1986, he was admitted to Purdue's *Book of Great Teachers* in 2008. He has published over a hundred papers in graph theory on a wide variety of topics, has served as Editor of the *College Mathematics Journal*, and has co-edited eight graph theory books with Robin Wilson.

Norman Biggs is Emeritus Professor of Mathematics at the London School of Economics. He has written many research papers and books on mathematics, including *Algebraic Graph Theory*, *Discrete Mathematics*, *Mathematics for Economics and Finance*, and *Codes: An Introduction to Information, Communication and Cryptography*. He has also written regularly on topics in the history of mathematics and numismatics. Since his 'retirement' in 2006 he has been teaching a Masters course at the LSE, and is currently setting up an undergraduate course on 'The history of mathematics in finance and economics'.

Andrea Bréard is an Associate Professor in the Mathematics Department at Lille University of Science and Technology, and Vice-President of the Humanities and Social Science Department of the École Polytechnique where she is also a part-time Professor of the History and Epistemology of Science. She received her PhD degree in Epistemology and the History of Science from TU Berlin and the Université Paris Diderot in 1997. Her thesis, *Re-Creation of a Mathematical Concept in Chinese Discourse: Series from the 1st to the 19th Century*, was published in 1999 in German. Recently she completed a book manuscript on *Reform, Bureaucratic Expansion and Production of Numbers: Statistics in China at the Turn of the 20th Century* and currently works as a Research Scholar at the IKGF in Erlangen, analysing the interrelation between games of chance, divination, and mathematical knowledge and rationality, as described in late Imperial Chinese sources.

Peter J. Cameron is Professor of Mathematics at Queen Mary, University of London, where he has been since 1986, following a position as tutorial fellow at Merton College, Oxford. Since his DPhil in Oxford, he has been interested in a variety of topics in algebra and combinatorics, especially their interactions. He has held visiting positions at the University of Michigan, California Institute of Technology, and the University of Sydney. He is currently chair of the British Combinatorial Committee.

Ahmed Djebbar is an Emeritus Professor of the History of Mathematics at Lille University of Science and Technology. He researches in the history of mathematics, specializing in the mathematical activities of the Islamic West (Andalus, Maghreb, and sub-Saharan Africa). He has published more than 150 research articles on the history of mathematics in Islamic countries, and has also published, on his own and in collaboration with other researchers, a score of works on the history of scientific activities in Islamic countries and in Africa. Since 1986 he has been Secretary of the African Mathematical Union's Commission on the History of Mathematics.

A. W. F. Edwards is a statistician, geneticist, and evolutionary biologist. He is a Fellow of Gonville and Caius College and retired Professor of Biometry at the University of

Cambridge. A pupil of R. A. Fisher's, he is best known for his pioneering work with L. L. Cavalli-Sforza on quantitative methods of phylogenetic analysis, and for advocating Fisher's concept of likelihood as the basis for statistical inference. He has written extensively on the history of genetics and statistics, including an analysis of Mendel's results, and on mathematical subjects such as Venn diagrams and Faulhaber polynomials. His books include *Likelihood*, *Foundations of Mathematical Genetics*, *Pascal's Arithmetical Triangle*, and *Cogwheels of the Mind: The Story of Venn Diagrams*.

Victor J. Katz is an Emeritus Professor of Mathematics at the University of the District of Columbia. The third edition of his *A History of Mathematics: An Introduction*, appeared in 2008. He is also the editor of *The Mathematics of Egypt, Mesopotamia, China, India and Islam: A Sourcebook*, as well as three Mathematical Association of America (MAA) books dealing with the use of the history of mathematics in teaching and two collections of historical articles taken from MAA journals. He directed two projects that helped college and secondary teachers learn the history of mathematics and its use in teaching, and was also the founding editor of *Loci: Convergence*, the MAA's online history of mathematics magazine.

Eberhard Knobloch is an Emeritus Professor of the History of Science and Technology at the Berlin University of Technology, an Academy Professor at the Berlin–Brandenburg Academy of Sciences (the former Prussian Academy of Sciences), and an Honorary Professor at the Chinese Academy of Sciences in Beijing. He has written or edited about 300 papers and books on the history and philosophy of the mathematical sciences, and on Renaissance technology, especially dealing with Archimedes, Leibniz, Alexander von Humboldt, and Borel. He was chief editor of *Historia Mathematica* and Chairman of the International Commission on the History of Mathematics. He is President of the International Academy of Sciences, Paris, and a past President of the European Society for the History of Science.

Donald E. Knuth is an Emeritus Professor at Stanford University, where he has worked with students in the computer science department since 1968, and he is also a Visiting Professor in Computer Science at Oxford University. Among his books are *The Art of Computer Programming* (four volumes so far), *Computers and Typesetting* (five volumes), *Concrete Mathematics*, *3:16 Bible Texts Illuminated*, and nine volumes of collected papers. His software systems TeX and METAFONT are used for the majority of today's mathematical publications and have more than a million users worldwide.

Takanori Kusuba is Professor of History and Philosophy of Science at Osaka University of Economics. In his PhD thesis, submitted to Brown University in 1993, he published a critical edition with English translation and commentary of the last two chapters (on combinatorics and magic squares) of the *Ganitakaumudī*, written by Nārayana Pandita in 1356, and has worked on the whole text. Jointly with David Pingree, he published *Arabic Astronomy in Sanskrit* in 2002, and has published several articles on Indian and Arabic mathematics. Recent publications include a critical edition (with English translation and mathematical commentary) of the *Ganitas ārakaumudī*, a middle-Indic mathematical text composed in the early 14th century.

E. Keith Lloyd spent the whole of his mathematical career at the University of Southampton, latterly as a senior lecturer. His main research interests were in combinatorics, including its history, and in mathematical chemistry. He is a co-author (with Norman L. Biggs and Robin J. Wilson) of *Graph Theory 1736–1936* and has other publications on graph theory and combinatorics. He has also published biographical articles on J. Howard Redfield (a pioneer in using group theory in enumeration) and on Peter Nicholson (an architect and mathematician). He has made several visits to China to deliver lectures on combinatorics.

Kim Plofker is a Visiting Assistant Professor in the Department of Mathematics at Union College in Schenectady, New York. She received her PhD degree from the Department of the History of Mathematics at Brown University under the late David Pingree. Her areas of interest include the history of mathematics and astronomy in India and the medieval Islamic world, and the cross-cultural transmission of scientific models. Recent publications include ' "Yavana" and "Indian": Transmission and foreign identity in the exact sciences' in *Annals of Science*, the survey book *Mathematics in India* in 2009, and 'Euler and Indian astronomy' in *Leonhard Euler: Life, Work and Legacy*.

John J. Watkins is an Emeritus Professor of Mathematics at Colorado College, USA. He received his PhD degree from the University of Kansas, specializing in commutative algebra. His main research interest, however, has been in graph theory, and he has published mainly in this area, including many papers with undergraduates as co-authors. He has written several books, including *Graphs: An Introductory Approach* (with Robin J. Wilson), *Across the Board: The Mathematics of Chessboard Problems*, and *Topics in Commutative Ring Theory*, and has recently finished his latest book, *Number Theory: A Historical Approach*.

Robin Wilson is an Emeritus Professor of Pure Mathematics at the Open University, Emeritus Professor of Geometry at Gresham College, London, and a former Fellow of Keble College, Oxford. He is currently President of the British Society for the History of Mathematics. He has written and edited many books on graph theory, including *Introduction to Graph Theory* and *Four Colours Suffice*, and on the history of mathematics, including *Lewis Carroll in Numberland*. He is involved with the popularization and communication of mathematics and its history, and was awarded the Mathematical Association of America's Lester Ford and Pólya prizes for 'outstanding expository writing'.

PICTURE SOURCES AND ACKNOWLEDGEMENTS

Frontispiece

A. Kircher, *Ars Magna Sciendi Sive Combinatoria,* Johannes Janssonius a Waesberge, Amsterdam (1669), frontispiece.

Introduction

This chapter is adapted, with permission, from D. E. Knuth, *The Art of Computer Programming*, Vol. 4A, *Combinatorial Algorithms*, Part 1, Addison-Wesley (2011), Section 7.2.1.7.

[2] A. Kircher, *Ars Magna Sciendi Sive Combinatoria*, Johannes Janssonius a Waesberge, Amsterdam (1669), 170.
[8] A. Kircher, *Musurgia Universalis* 2, Haeredes Francisci Corbelletti, Rome (1650), 32.
[9] S. Donnolo, Hakhmoni (*Commentary on Sefer Yetzirah*), 946 (reprinted in Warsaw, 1884).
[10] M. Mersenne, Traitez de la voix et des chants, *Harmonie Universelle*, Vol. 2, Sebastien Cramoisy, Paris (1636), 117–28.
[10] A. Kircher, *Musurgia Universalis* 2, Haeredes Francisci Corbelletti, Rome (1650), 11–12.
[15] R. Llull, *Ars Demonstrativa*, Biblioteca Marciana, VI 200, f. 3^{v}.
[16] Codex St Peter perg 92, ff. 28^{v}, 39^{v}, Badische Landesbibliothek, Karlsruhe.
[25] From a copy in the collection of Tamaki Yano at Saitama University, Japan.
[25] *Encyclopedia Japonicae*, Sanseido, Tokyo (1910), 1299.
[26] Anonymous (attributed to George and/or Richard Puttenham), *The Art of English Poesie*, Richard Field (1589), 70–2.
[32] A. Cayley, On the theory of the analytical forms called trees. Second part, *Philos. Mag.* (4) 18 (1859), 374–8 = *Collected Papers* 4, 112–15.

Chapter 1

[40] Statue of Viṣṇu (Vishnu) from S. G. Goodrich, *Lights and Shadows of Asiatic History*, Bradbury, Soden & Co., Boston (1844).

[47] Photograph of Piṅgala's fountain courtesy of Raazesh Sainudiin, who co-designed the fountain with Krithika Yogeeswaran.

[49/53] Al-Bīrūnī, *Kitab fi tahqiq-i-ma li'l-Hind* [*India*], 11th century, Arabic edition, Hyderabad, pp. 110–11.

[57] Nepalese Sanskrit manuscript of Nārāyaṇa's *Gaṇitakaumudī*, ff. 40a, 36b, 32b, 18b.

Chapter 2

[64] Zhu Shijie, *Siyuan yujian* (1303), frontispiece (reprinted by Shangwu yinshuguan, Shanghai, 1937).

[66] Yu Xiangdou (ed.), *Santai wanyong zhengzong* (Santai's Orthodox Instructions for Myriad Users) (1599).

[67] Cheng Dawei, *Suanfa tongzong* (1592) (reprinted by Taishantung, Jingdu, 1882).

[68] Yang Hui, *Xugu zhaiqi suanfa* (1275), diagrams courtesy of Andrea Bréard.

[69] Bao Qishou, *Zengbu suanfa hunyuantu*, from his collected writings *Binai shanfang ji* (*c.*1880).

[74/76] Wang Lai, *Dijian shuli*, in *Hengzhai suanxue* (Hengzhai's Mathematics), Vol. 4, Jiashutang, China (1854), 8a–8b and 9a–9b.

[79] Li Shanlan, *Duoji bilei*, 4 juan, in *Zeguxizhai suanxue*, Vol. 4, scroll 3, Jinling ed. (1867), 1a–1b.

Chapter 3

[82] Al-Karajī (*c.*1007), as discussed in later algebraic writings of as-Samaw'al.

[85] Ibn Durayd, *Jamshara*, drawing courtesy of Ahmed Djebbar.

[86] Circles of al-Khalīl, anonymous manuscript (*c.*770); see as-Suyūtī, *al-Muzhir fīᶜūlum al-lugha* (book), Cairo (undated) 71–6.

[91] Naṣīr ad-Dīn aṭ-Ṭūsī, *Kitāb jawāmiᶜ al-hisāb bi t-takht wa-t-turāb* (*c.*1250).

[96] Ibn Munᶜim, *Fiqh al-hisāb*, 328; the manuscript containing the arithmetical triangle was copied in the 15th century.

Chapter 4

[108] *Sefer Yetsirah* (origin unknown), title page; this work exists in multiple versions of varying lengths.

[115] Colophon from the Moscow 1093 ms of *Maasei Hoshev*, second edition (1322).

Chapter 5

[122] M. Mersenne, *Harmonie Universelle*, Vol. 2, Sebastien Cramoisy, Paris (1636), title page.

[124] R. Llull, *Ars Magna*, in *Opera ea Quae ad Adinventam ab ipso Artem Universalem Scientiarum Artiumque Omnium Brevi Compendio . . . Pertinent*, Strassburg (1598).

[126] 'Pseudo-Ovidius' (R. de Fournival?), *De Vetula* (*c.*1250), verses 420–41; the work was published in 1470.

[127] This 1553 image of G. Cardano appears in his *De Rerum Varietate, Libri XVIII* (1557).
[130] Image of C. Clavius courtesy of Staatsbibliothek Bamberg VA 134d (photo Gerald Raab).
[131] C. Clavius, *In Sphaeram Ioannis de Sacro Bosco Commentarius*, Rome (1570); last (seventh) version in *Opera Mathematica*, Vol. 3, 1st part, Anton Hierat, Mainz (1611), 17.
[133] M. Mersenne, *Harmonicorum Libri*, Guillaume Baudry, Paris (1635), 117.
[135] A. Kircher in 1664, from his *Romani Collegii Societatis Jesu Musaeum Celeberrimum*, Amsterdam (1678).
[140] J. Caramuel, *Mathesis Biceps Vetus et Nova*, Campagna (1670), title page.

Chapter 6

[146] G. W. Leibniz's *Dissertatio de Arte Combinatoria*, Fick & Seubold, Leipzig (1666), title page = *Mathematische Schriften*, Vol. 5, 7.
[149] 19th-century engraving of G. W. Leibniz by B. Holl, published by W. S. Orr & Co., London.
[156] 18th-century engraving of Jacob Bernoulli by P. Dupin.
[158] Jacob Bernoulli, *Ars Conjectandi*, Brothers Thurneisen, Basel (1713), 97 = *Die Werke*, Vol. 3, Naturforschende Gesellschaft, Birkhäuser (1975), 166.
[159] P. R. de Montmort, *Essay d'Analyse sur les Jeux de Hazard*, 2nd edn., Quillau, Paris (1713), title page (reprinted by Chelsea, 1980).

Chapter 7

[166] B. Pascal, *Traité du Triangle Arithmétique*, Desprez, Paris (1665), frontispiece.
[172] N. Tartaglia, *General Trattato di Numeri et Misure*, Part II, Vinegia (1556), ff. 69r–73v.
[172] M. Stifel, *Arithmetica Integra*, Nuremberg (1544), ff. 44v, 45v.
[173] G. Cardano, *Opus Novum de Proportionibus Numerorum*, Basilea (1570), 185.
[173] M. Mersenne, *Harmonicorum Libri XII*, Lutetiae Parisiorum (1636), 136.
[174] 17th-century engraving of B. Pascal by G. Edelinck.
[175] B. Pascal, *Traité du Triangle Arithmétique*, Desprez, Paris (1665), title page.
[177] Jacob Bernoulli, *Ars Conjectandi*, Brothers Thurneisen, Basel (1713), title page.

Chapter 8

[182] J. J. Sylvester, On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, with three appendices, *Amer. J. Math.* 1 (1878), 64–125, on p. 109 = *Math. Papers*, Vol. 3, 148–206, on p. 190.
[184] M. Zeiller, *Topographia Prussiae et Pomerelliae*, Frankfurt (1652).
[184] L. Euler, Solutio problematis ad geometriam situs pertinentis, *Comm. Acad. Sci. Imp. Petropol.* 8 (1736), 128–40, on p. 129 = *Opera Omnia* (1), Vol. 7, 1–10.
[186] T. Clausen, [Second postscript to] De linearum tertii ordinis proprietatibus, *Astron. Nachr.* 21 (1844), col. 209–16.
[187] J. B. Listing, *Vorstudien zur Topologie*, Göttinger Studien (Abt. 1), *Math. Naturwiss.* Abh. 1 (1847), 811–75 = separately as book, Göttingen (1848).

[191] W. R. Hamilton, [Instructions for the Icosian Game], John Jaques and Son, London (1859).
[192] L. Euler, *Opera Omnia* (4), I, letter 863.
[197] W. K. Clifford, *Mathematical Fragments, being Facsimiles of his Unfinished Papers relating to the Theory of Graphs*, Macmillan, London (1881), collection of Robin Wilson.
[201] Photograph, collection of Robin Wilson; the map appeared in P. J. Heawood, Map-colour theorem, *Quart. J. Pure Appl. Math.* 24 (1890), 332–8.

Chapter 9

[204] L. Euler, engraving by Bartholomaeu, Hübner (1786) (signed also by Christian v. Mechel).
[209/219/224] Photographs of Sylvester, Hardy, and MacMahon courtesy of the London Mathematical Society.
[219] G. H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, *Proc. London Math. Soc. (2)* 17 (1918), 75–115, on pp. 84–5 and 114–15 = G. H. Hardy, *Math. Papers*, Vol. 1, pp. 306–39, on pp. 284–5 and 338.

Chapter 10

[230] Photograph of Kirkman courtesy of the London Mathematical Society.
[237] *The Lady's and Gentleman's Diary* (1850), title page.
[244] Copyright photograph of Fisher by A. Barrington Brown, reproduced by permission of the Fisher Memorial Trust.
[244] Photograph of Yates courtesy of the Rothamsted Archives.

Chapter 11

[250] G. Perec, *La Vie Mode d'Emploi*, Hachette (1978), cover design, reprinted with the publisher's permission.
[251] R. Llull, *Ars Magna*, in *Opera ea Quae ad Adinventam ab ipso Artem Universalem Scientiarum Artuiumque Omnium Brevi Compendio . . . Pertinent*, Strassburg (1598), and Al-Būnī, *Shams al-Ma'arif al-Kubra* (*c.*1200).
[252] From D. B. MacDonald, Description of a silver amulet, *Z. Assyriologie Verwandte Gebiete* 26 (1912), 267–9 and S. Seligmann, Das Siebenschläfer-Amulet, *Der Islam* 5 (1914), 370–88.
[253] J. Shurreef, *Qanoon-E-Islam* (1832), transl. by G. A. Herklotz as *Customs of the Moosulmans of India; Comprising a Full and Exact Account of their Various Rites and ceremonies from the Moment of Birth until Death*.
[254] Al-Būnī, *Shams al-Ma'arif al-Kubra* (*c.*1200).
[255] C. G. Bachet de Méziriac, *Problèmes Plaisans et Délectables qui se font par les Nombres* (1612), and J. H. van Lint and R. M. Wilson, *A Course in Combinatorics* (1992), cover design, courtesy of Cambridge University Press.
[256] F. C. Cretté de Palluel, On the advantage and economy of feeding sheep in the house with roots, *Ann. Agric.* 14 (1790), 51–5.
[257] Photograph courtesy of A. W. F. Edwards.

[258/259] L. Euler, Recherches sur une nouvelle espèce de quarrés magiques, *Verh. Zeeuwsch. Genootsch. Wetensch. Vlissingen* 9 (1782), 85–239, on pp. 85–6, 89, and 91 = *Opera Omnia* (1), Vol. 7 (1792), 291–392, on pp. 291, 294.

[266] J. Sauveur, Construction générale des quarrés magiques, *Mém. Acad. Royales Sciences* (1710), 92–138.

[268] M. Gardner, Euler's spoilers: The discovery of an order-10 Graeco-Latin square, *New Mathematical Diversions*, Mathematical Association of America (1995).

[277/278/279] C. Boyer, Sudoku's French ancestors, *Math. Intelligencer* 29(1) (2007), 37–44 and 29(2) (2007), 59–63, courtesy of *The Mathematical Intelligencer*.

Chapter 12

[284] P. Nicholson, *Practical Carpentry, Joinery, and Cabinet-Making* . . . (revised by Thomas Tredgold), Thomas Kelly, London (1852), frontispiece.

[289] C. F. Hindenburg, *Der polynomische Lehrsatz*, Leipzig (1796), title page and p. 298.

[290] P. Nicholson, *Essays on the Combinatorial Analysis*, London (1818), title page.

[293] A. Cayley, *Illustrated London News* (15 September 1883).

[300] J. H. Redfield, The theory of group-reduced distributions, *Amer. J. Math.* 49 (1927), 432–55, on p. 443.

[302] Photograph of Redfield by kind permission of the late Mrs Priscilla Redfield Roe (daughter of J. H. Redfield).

[302] Photograph of Pólya from the collection of G. Alexanderson.

Chapter 13

[308/313] A. de Moivre, *The Doctrine of Chances*, title page and p. 96.

[317/321] Photographs of Ramsey and Hall courtesy of the London Mathematical Society.

[317] Photograph of Erdős from the collection of G. Alexanderson.

[321] Photograph of König courtesy of the János Bolyai Mathematical Society.

Chapter 14

[330] A page of Appel and Haken's reducible configurations, courtesy of Kenneth Appel and Wolfgang Haken.

[332] H. E. Dudeney, *Amusements in Mathematics*, Nelson, London (1917), 73.

[333] G. D. Birkhoff, The reducibility of maps, *Amer. J. Math.* 35 (1913), 115–28.

[336] Photograph of Appel and Haken courtesy of the University of Illinois.

[345] Photograph of Petersen courtesy of the Danish Mathematical Society.

Aftermath

[354] Photograph by Knut Falch courtesy of the Norwegian Academy of Science and Letters.

This page intentionally left blank

INDEX