Sudoku: Is it Mathematics?

Peter J. Cameron

Forder lectures
April 2008

There's no mathematics involved. Use logic and reasoning to solve the puzzle.

Instructions in *The Independent*

There's no mathematics involved. Use logic and reasoning to solve the puzzle.

Instructions in *The Independent*

Mathematics $\neq$ reasoning and logic???

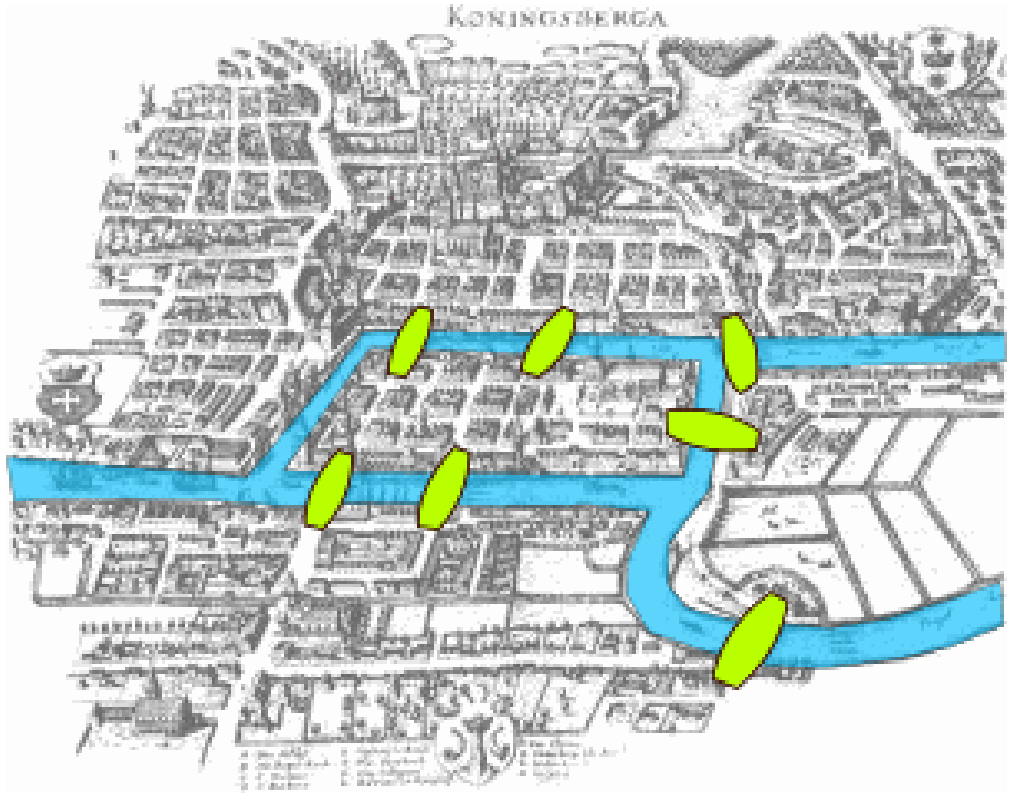
Technology transfer

To criticize mathematics for its abstraction is to miss the point entirely. Abstraction is what makes mathematics work. If you concentrate too closely on too limited an application of a mathematical idea, you rob the mathematician of his most important tools: analogy, generality, and simplicity. Mathematics is the ultimate in technology transfer.

Ian Stewart, *Does God play dice? The mathematics of chaos*, Penguin, London, 1990.

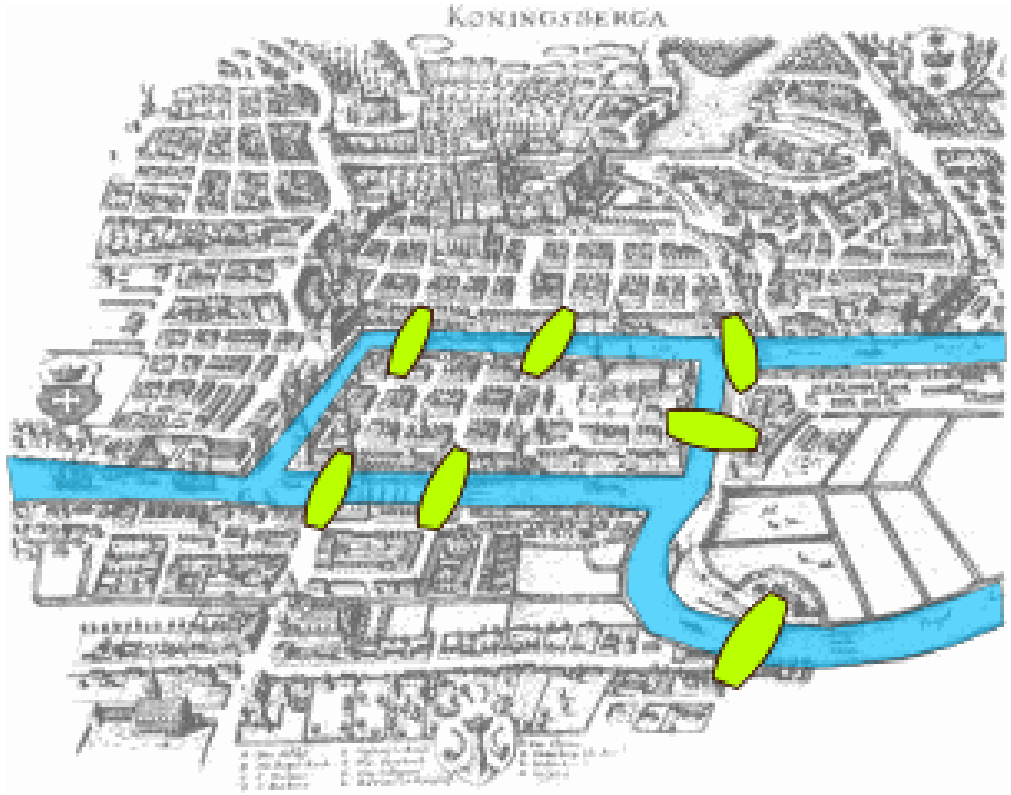
Euler

The bridges of Königsberg



Is it possible to walk around the town, crossing each bridge exactly once?

The bridges of Königsberg



Is it possible to walk around the town, crossing each bridge exactly once?
Euler showed: No!

What is mathematics?

Leonhard Euler, Letter to Carl Ehler, mayor of Danzig, 3 April 1736:

What is mathematics?

Leonhard Euler, Letter to Carl Ehler, mayor of Danzig, 3 April 1736:

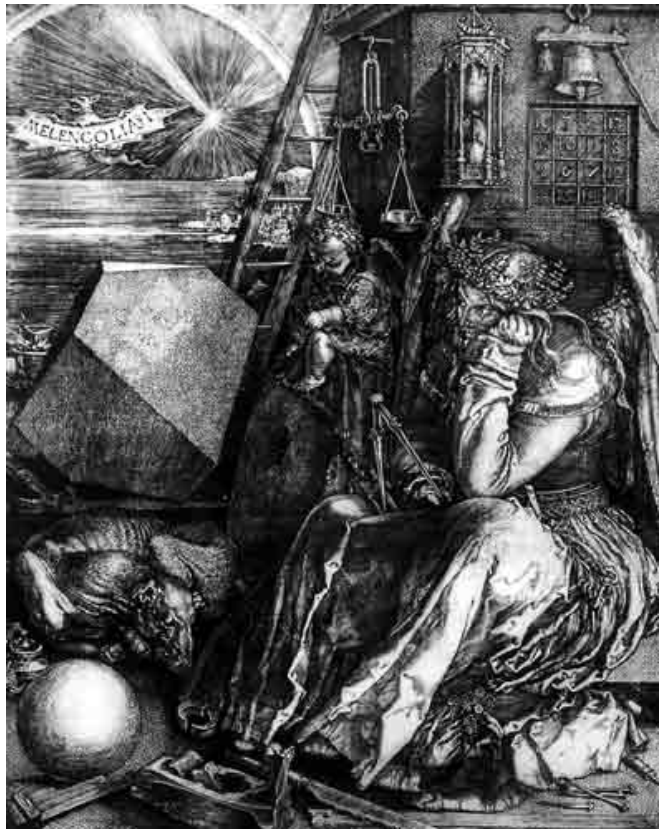
> *Thus you see, most noble Sir, how this type of solution [to the Königsberg bridge problem] bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle . . .*

What is mathematics?

Leonhard Euler, Letter to Carl Ehler, mayor of Danzig, 3 April 1736:

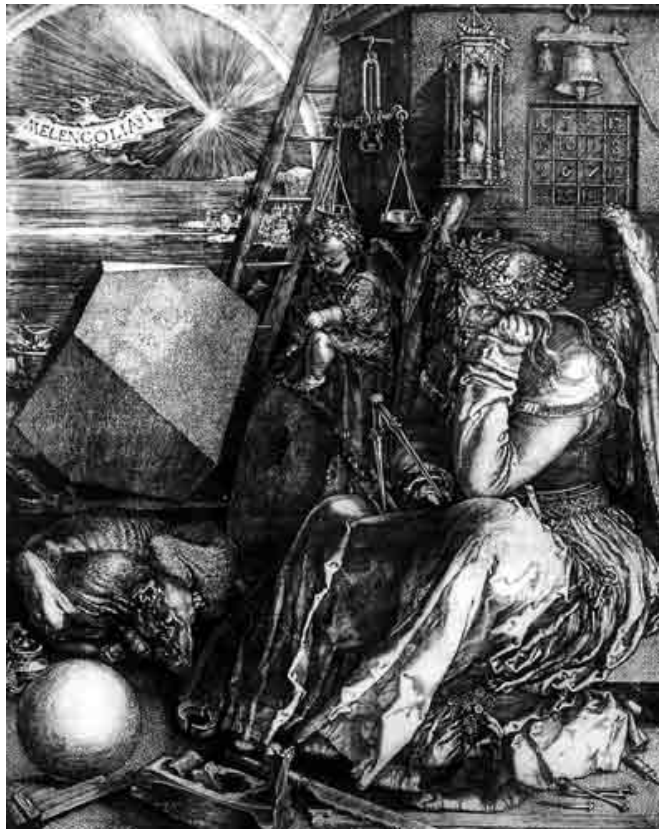
> *Thus you see, most noble Sir, how this type of solution [to the Königsberg bridge problem] bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle . . .*
>
> *In the meantime, most noble Sir, you have assigned this question to the geometry of position, but I am ignorant as to what this new discipline involves, and as to which types of problem Leibniz and Wolff expected to see expressed in this way.*

Dürer's *Melancholia*



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

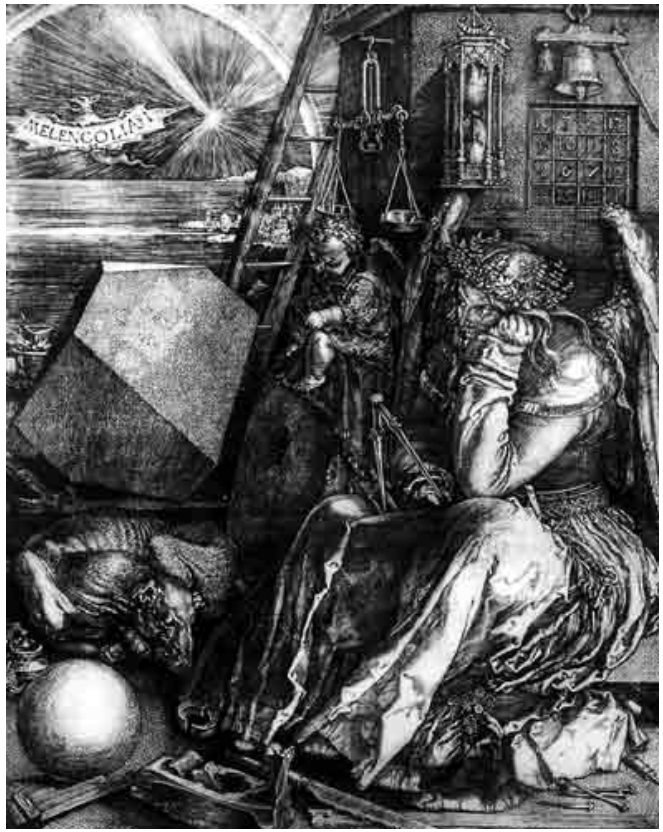
Dürer's *Melancholia*



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

All rows, columns, and diagonals sum to 34.

Dürer's *Melancholia*



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

All rows, columns, and diagonals sum to 34. The date of the picture is included in the square.

Euler's construction

Take a Graeco-Latin square of order n.

$C\beta$	$A\gamma$	$B\alpha$
$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$

Euler's construction

Take a Graeco-Latin square of order n. Replace the symbols by $0, 1, \ldots, n-1$.

$C\beta$	$A\gamma$	$B\alpha$
$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$

21	02	10
00	11	22
12	20	01

Euler's construction

Take a Graeco-Latin square of order n. Replace the symbols by $0, 1, \ldots, n-1$. Interpret the result as a two-digit number in base n. Add one.

$C\beta$	$A\gamma$	$B\alpha$
$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$

21	02	10
00	11	22
12	20	01

8	3	4
1	5	9
6	7	2

Euler's construction

Take a Graeco-Latin square of order n. Replace the symbols by $0, 1, \ldots, n-1$. Interpret the result as a two-digit number in base n. Add one.

$C\beta$	$A\gamma$	$B\alpha$
$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$

21	02	10
00	11	22
12	20	01

8	3	4
1	5	9
6	7	2

Some rearrangement may be needed to get the diagonal sums correct.

Euler's construction

Take a Graeco-Latin square of order n. Replace the symbols by $0, 1, \ldots, n-1$. Interpret the result as a two-digit number in base n. Add one.

$C\beta$	$A\gamma$	$B\alpha$
$A\alpha$	$B\beta$	$C\gamma$
$B\gamma$	$C\alpha$	$A\beta$

21	02	10
00	11	22
12	20	01

8	3	4
1	5	9
6	7	2

Some rearrangement may be needed to get the diagonal sums correct.

So for which n do Graeco-Latin squares exist?

Euler's officers

Six different regiments have six officers, each one holding a different rank (of six different ranks altogether). Can these 36 officers be arranged in a square formation so that each row and column contains one officer of each rank and one from each regiment?

Euler's officers

> *Six different regiments have six officers, each one holding a different rank (of six different ranks altogether). Can these 36 officers be arranged in a square formation so that each row and column contains one officer of each rank and one from each regiment?*

Trial and error suggests the answer is "No":

Euler's conjeture

Euler knew how to construct a Graeco-Latin square of every order n not congruent to 2 mod 4.

Euler's conjeture

Euler knew how to construct a Graeco-Latin square of every order n not congruent to 2 mod 4.

It is trivial that there is no Graeco-Latin square of order 2.

Euler's conjeture

Euler knew how to construct a Graeco-Latin square of every order n not congruent to 2 mod 4.

It is trivial that there is no Graeco-Latin square of order 2.

In 1900, Tarry confirmed that there is no Graeco-Latin square of order 6.

Euler's conjeture

Euler knew how to construct a Graeco-Latin square of every order n not congruent to 2 mod 4.

It is trivial that there is no Graeco-Latin square of order 2.

In 1900, Tarry confirmed that there is no Graeco-Latin square of order 6.

In 1960, Bose, Shrikhande and Parker showed that, apart from these two cases, Euler was wrong: Graeco-Latin squares exist for all other orders.

Latin squares

A Latin square is the type of structure formed by the Latin letters in a Graeco-Latin square: that is, each symbol occurs exactly once in each row or column.

Latin squares

A Latin square is the type of structure formed by the Latin letters in a Graeco-Latin square: that is, each symbol occurs exactly once in each row or column.

There is no question about the existence of Latin squares: there is a Latin square of any order. But we still don't know many things about them, for example, how many there are.

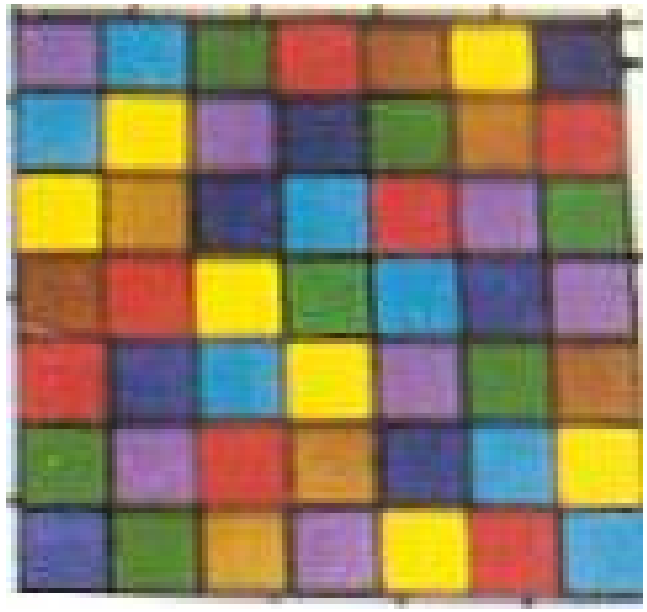
Latin squares

A Latin square is the type of structure formed by the Latin letters in a Graeco-Latin square: that is, each symbol occurs exactly once in each row or column.

There is no question about the existence of Latin squares: there is a Latin square of any order. But we still don't know many things about them, for example, how many there are.

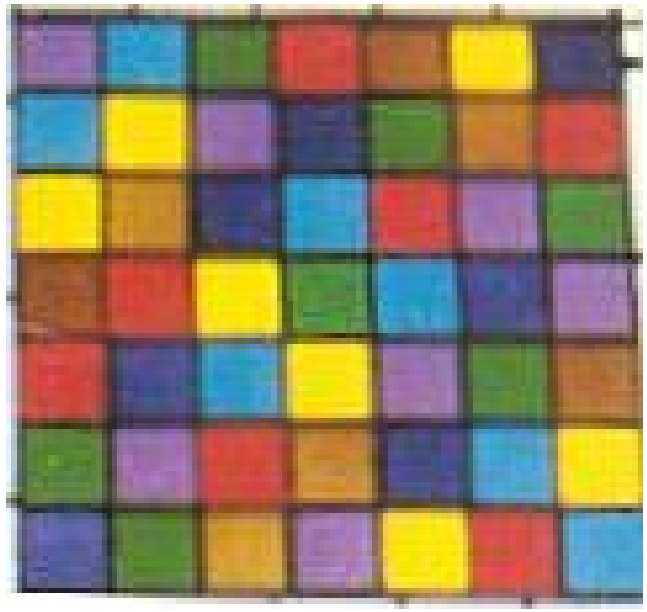
We also don't know whether there is an efficient way to decide if a given Latin square can be extended to a Graeco-Latin square.

Latin squares in statistics



Latin squares were introduced into statistics by R. A. Fisher.

Latin squares in statistics



Latin squares were introduced into statistics by R. A. Fisher.

They are useful for design of experiments in field trials where there may be spatial effects.

Latin squares in statistics

A Latin square at Rothamsted Experimental Station.

This Latin square was designed by Rosemary Bailey. Thanks to Sue Welham for the photograph.

Latin squares in statistics

A Latin square at Rothamsted Experimental Station.

This Latin square was designed by Rosemary Bailey. Thanks to Sue Welham for the photograph.

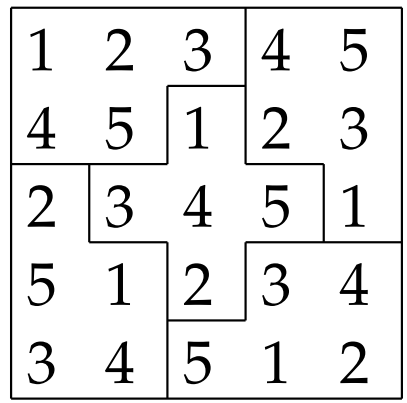
It has the additional property of being complete: each ordered pair of distinct symbols occurs together once in a row and once in a column.

Gerechte designs

W. Behrens: What if there is, for example, a boggy patch in the middle of the field?

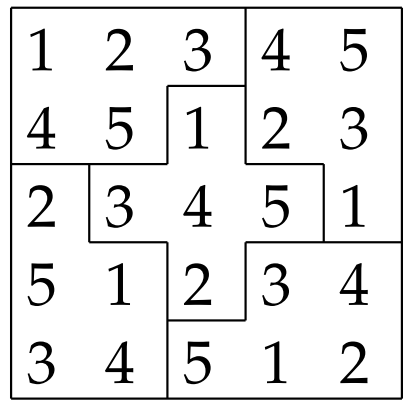
Gerechte designs

W. Behrens: What if there is, for example, a boggy patch in the middle of the field?



Gerechte designs

W. Behrens: What if there is, for example, a boggy patch in the middle of the field?



This is a gerechte design (a “fair design”).

Critical sets

John Nelder: A critical set is a partially filled Latin square which can be completed in a unique way to a Latin square, but if any entry is deleted the completion is no longer unique.

1	2		
2			
			3

Critical sets

John Nelder: A critical set is a partially filled Latin square which can be completed in a unique way to a Latin square, but if any entry is deleted the completion is no longer unique.

1	2		
2			
			3

Critical sets were designed to study the process of "stepping" between different Latin squares by means of *trades*.

Sudoku

So a Sudoku puzzle is a partial gerechte design for the partition of a 9×9 square into nine 3×3 subsquares, which contains a critical set.

Sudoku

So a Sudoku puzzle is a partial gerechte design for the partition of a 9×9 square into nine 3×3 subsquares, which contains a critical set.

In fact Sudoku was invented by Howard Garns (a retired New York architect) in the 1980s, under the name "number place".

Sudoku

So a Sudoku puzzle is a partial gerechte design for the partition of a 9×9 square into nine 3×3 subsquares, which contains a critical set.

In fact Sudoku was invented by Howard Garns (a retired New York architect) in the 1980s, under the name "number place".

It was popularised in Japan by Maki Kaji, who renamed it *Su Doku*.

Sudoku

So a Sudoku puzzle is a partial gerechte design for the partition of a 9×9 square into nine 3×3 subsquares, which contains a critical set.

In fact Sudoku was invented by Howard Garns (a retired New York architect) in the 1980s, under the name "number place".

It was popularised in Japan by Maki Kaji, who renamed it *Su Doku*.

New Zealander Wayne Gould popularised it in the West. The rest is history...

How many Sudoku solutions?

Felgenhauer and Jarvis, showed, by a massive computation, that the number of different Sudoku solutions (filled Sudoku grids) is

How many Sudoku solutions?

Felgenhauer and Jarvis, showed, by a massive computation, that the number of different Sudoku solutions (filled Sudoku grids) is

$$6\,670\,903\,752\,021\,072\,936\,960.$$

How many Sudoku solutions?

Felgenhauer and Jarvis, showed, by a massive computation, that the number of different Sudoku solutions (filled Sudoku grids) is

$$6\,670\,903\,752\,021\,072\,936\,960.$$

This figure has been independently verified.

How many Sudoku solutions?

We count Sudoku solutions up to

- Permuting the numbers $1, \ldots, 9$;
- Permuting rows and columns preserving the partitions into 3 sets of 3;
- Possibly transposing the grid.

How many Sudoku solutions?

We count Sudoku solutions up to

- Permuting the numbers $1, \ldots, 9$;
- Permuting rows and columns preserving the partitions into 3 sets of 3;
- Possibly transposing the grid.

The number of different solutions of ordinary Sudoku (with these rules) is 5 472 730 538.

How many Sudoku solutions?

We count Sudoku solutions up to

- Permuting the numbers $1, \ldots, 9$;
- Permuting rows and columns preserving the partitions into 3 sets of 3;
- Possibly transposing the grid.

The number of different solutions of ordinary Sudoku (with these rules) is 5 472 730 538.
This was computed by Jarvis and Russell using the Orbit-counting Lemma applied to the group $S_9 \times ((S_3 \operatorname{wr} S_3) \operatorname{wr} S_2)$ of order $9! \cdot 6^8 \cdot 2$ acting on the set of solutions counted by Felgenhauer and Jervis.

Symmetric Sudoku

This was invented by Robert Connelly and independently by Vaughan Jones. It s connected to some very interesting and important mathematical topics.

Symmetric Sudoku

This was invented by Robert Connelly and independently by Vaughan Jones. It s connected to some very interesting and important mathematical topics.

Each number from 1 to 9 should occur once in each set of the following types:

- rows;
- columns;
- 3×3 subsquares;
- broken rows (one of these consists of three "short rows" in the same position in the three subsquares in a large column);
- broken columns (similarly defined);
- locations (a location consists of the nine cells in a given position, e.g. middle of bottom row, in each of the nine subsquares).

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Rows

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Columns

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Subsquares

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Broken rows

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Broken columns

Example

3	5	9	2	4	8	1	6	7
4	8	1	6	7	3	5	9	2
7	2	6	9	1	5	8	3	4
8	1	4	7	3	6	9	2	5
2	6	7	1	5	9	3	4	8
5	9	3	4	8	2	6	7	1
6	7	2	5	9	1	4	8	3
9	3	5	8	2	4	7	1	6
1	4	8	3	6	7	2	5	9

Locations

Affine geometry

We coordinatise the cells of the grid with F^4, where F is the integers mod 3, as follows:

- the first coordinate labels large rows;
- the second coordinate labels small rows within large rows;
- the third coordinate labels large columns;
- the fourth coordinate labels small columns within large columns.

Affine geometry

We coordinatise the cells of the grid with F^4, where F is the integers mod 3, as follows:

- the first coordinate labels large rows;
- the second coordinate labels small rows within large rows;
- the third coordinate labels large columns;
- the fourth coordinate labels small columns within large columns.

Now the relevant regions are cosets of the following subspaces:

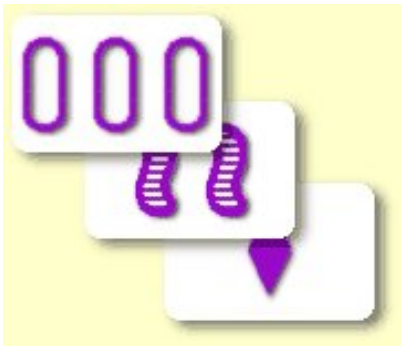
Rows	$x_1 = x_2 = 0$	Columns	$x_3 = x_4 = 0$
Subsquares	$x_1 = x_3 = 0$	Broken rows	$x_2 = x_3 = 0$
Broken columns	$x_1 = x_4 = 0$	Locations	$x_2 = x_4 = 0$

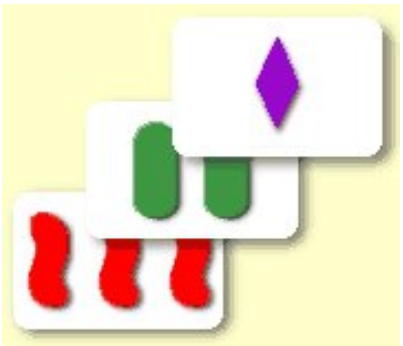
Affine spaces and SET

The card game SET has 81 cards, each of which has four attributes taking three possible values (number of symbols, shape, colour, and shading). A winning combination is a set of three cards on which either the attributes are all the same, or they are all different.

Affine spaces and SET

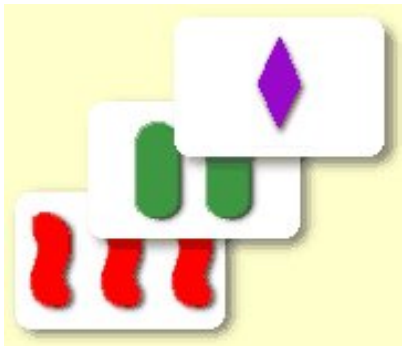
The card game SET has 81 cards, each of which has four attributes taking three possible values (number of symbols, shape, colour, and shading). A winning combination is a set of three cards on which either the attributes are all the same, or they are all different.





Affine spaces and SET

The card game SET has 81 cards, each of which has four attributes taking three possible values (number of symbols, shape, colour, and shading). A winning combination is a set of three cards on which either the attributes are all the same, or they are all different.



Each card has four coordinates taken from F (the integers mod 3), so the set of cards is identified with the 4-dimensional affine space. Then the winning combinations are precisely the affine lines!

Coding theory

Coding theory was invented in the 1950s by Shannon, Hamming and Golay to solve the problem of transmitting information accurately through a "noisy" channel, in which some symbols are randomly changed during transmission.

Coding theory

Coding theory was invented in the 1950s by Shannon, Hamming and Golay to solve the problem of transmitting information accurately through a "noisy" channel, in which some symbols are randomly changed during transmission.

We transmit "words", which are strings of symbols taken from a fixed alphabet (in practice the binary alphabet $\{0, 1\}$, though any alphabet could be used). The strategy is that, instead of transmitting all possible strings, we restrict our messages to those belonging to a suitable "code". Codewords should have the property that any two of them are so different that, even if we garble one a bit, it still resembles the original more closely than it resembles any other.

An example

Alphabet $\{0, 1, 2\}$.

An example

Alphabet $\{0,1,2\}$.

$$C = \left\{ \begin{array}{lll} 0000 & 1012 & 2021 \\ 0111 & 1120 & 2102 \\ 0222 & 1201 & 2210 \end{array} \right\}.$$

An example

Alphabet $\{0, 1, 2\}$.

$$C = \left\{ \begin{array}{lll} 0000 & 1012 & 2021 \\ 0111 & 1120 & 2102 \\ 0222 & 1201 & 2210 \end{array} \right\}.$$

Any two codewords have distance 3.

An example

Alphabet $\{0, 1, 2\}$.

$$C = \left\{ \begin{array}{lll} 0000 & 1012 & 2021 \\ 0111 & 1120 & 2102 \\ 0222 & 1201 & 2210 \end{array} \right\}.$$

Any two codewords have distance 3.

For example, to change 1012 into 0111 we have to change the first, second, and fourth symbols.

An example

Alphabet $\{0, 1, 2\}$.

$$C = \left\{ \begin{array}{lll} 0000 & 1012 & 2021 \\ 0111 & 1120 & 2102 \\ 0222 & 1201 & 2210 \end{array} \right\}.$$

Any two codewords have distance 3.

For example, to change 1012 into 0111 we have to change the first, second, and fourth symbols.

So the code will correct a single error.

An example

Alphabet $\{0,1,2\}$.

$$C = \left\{ \begin{array}{lll} 0000 & 1012 & 2021 \\ 0111 & 1120 & 2102 \\ 0222 & 1201 & 2210 \end{array} \right\}.$$

Any two codewords have distance 3.

For example, to change 1012 into 0111 we have to change the first, second, and fourth symbols.

So the code will correct a single error.

For example, the word 1221 is one step away from 1201 but at least two steps from any other codeword.

Perfect codes

A code is a set C of "words" or n-tuples over a fixed alphabet F. The Hamming distance between two words v, w is the number of coordinates where they differ; that is, the number of errors needed to change the transmitted word v into the received word w.

Perfect codes

A code is a set C of "words" or n-tuples over a fixed alphabet F. The Hamming distance between two words v, w is the number of coordinates where they differ; that is, the number of errors needed to change the transmitted word v into the received word w.

A code C is e-error-correcting if there is *at most* one word at distance e or less from any codeword. [Equivalently, any two codewords have distance at least $2e + 1$.] We say that C is perfect e-error-correcting if "at most" is replaced here by "exactly".

Perfect codes

A code is a set C of "words" or n-tuples over a fixed alphabet F. The Hamming distance between two words v, w is the number of coordinates where they differ; that is, the number of errors needed to change the transmitted word v into the received word w.

A code C is e-error-correcting if there is *at most* one word at distance e or less from any codeword. [Equivalently, any two codewords have distance at least $2e + 1$.] We say that C is perfect e-error-correcting if "at most" is replaced here by "exactly".

The example on the last slide is a perfect code.

Perfect codes and symmetric Sudoku

- The positions of any symbol in a symmetric Sudoku solution form a perfect code.

Perfect codes and symmetric Sudoku

- The positions of any symbol in a symmetric Sudoku solution form a perfect code.
- So the entire solution is a partition of the affine space into nine perfect codes.

Perfect codes and symmetric Sudoku

- The positions of any symbol in a symmetric Sudoku solution form a perfect code.
- So the entire solution is a partition of the affine space into nine perfect codes.
- Using the SET test, a perfect code is an affine subspace.

Perfect codes and symmetric Sudoku

- The positions of any symbol in a symmetric Sudoku solution form a perfect code.
- So the entire solution is a partition of the affine space into nine perfect codes.
- Using the SET test, a perfect code is an affine subspace.
- So there are only two different symmetric Sudoku solutions.

Perfect codes and symmetric Sudoku

- The positions of any symbol in a symmetric Sudoku solution form a perfect code.
- So the entire solution is a partition of the affine space into nine perfect codes.
- Using the SET test, a perfect code is an affine subspace.
- So there are only two different symmetric Sudoku solutions.

No one would doubt that this really is mathematics!

The two symmetric Sudoku solutions