Mathematics of Sudoku

Shreya Nama

Supervised by Andrew Booker
Level 6
20 Credit Points

March 12, 2025

Acknowledgement of Sources

For all ideas taken from other sources (books, articles, internet), the source of the ideas is mentioned in the main text and fully referenced at the end of the report.

All material which is quoted essentially word-for-word from other sources is given in quotation marks and referenced.

Pictures and diagrams copied from the internet or other sources are labelled with a reference to the web page or book, article etc.

Signed ______________________

Date 11/03/2025

Contents

1 Introduction **4**
- 1.1 What is Sudoku? . . . 4
- 1.2 Where does Sudoku originate? . . . 6
- 1.3 Why? . . . 7
- 1.4 What questions can we answer? . . . 7
- 1.5 Some basic facts . . . 7

2 Counting Shidoku **9**
- 2.1 Shidoku . . . 9
- 2.2 Distinct Shidoku grids . . . 9

3 Transformations **14**
- 3.1 Which transformations can be applied? . . . 14
- 3.2 Effects of Transformations . . . 15
- 3.3 Essentially Different Shidoku grids . . . 16

4 Counting grids **18**
- 4.1 Burnside's Lemma . . . 18
- 4.2 Symmetry Groups . . . 20
- 4.3 Conjugacy classes . . . 22
- 4.4 Graphical representation . . . 26

5 Solving Puzzles **30**
- 5.1 Counting Clues . . . 30
- 5.2 Solving Strategies . . . 32
  - 5.2.1 Backtracking Algorithms . . . 32
  - 5.2.2 The Blocking Effect . . . 33

1 Introduction

1.1 What is Sudoku?

Sudoku is a puzzle based on logic, that relies on deductive reasoning and a variety of solving strategies- which likely contributes to it's widespread popularity. Sudoku puzzles are often presented with varying numbers of filled in cells of initial clues or givens- which affect the level of difficulty. In section 5 we will explore the association between the number of initial clues and the complexity of the ensuing puzzle.

The classic case of this puzzle is a 9×9 grid filled with 3×3 sub-grids, where every one of its 81 cells must be filled with the numbers 1-9, without them repeating in each row, column and sub-grid. Throughout this paper, the term *Sudoku*, will always refer to the standard 9×9 variant, unless specified otherwise. We can observe that Sudoku grids are a subclass of Latin squares. Sudoku adds the additional constraint of subgrids. Any solution to a Sudoku puzzle is a Latin square, with additional restrictions imposed.

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

Figure 1: An easy level Sudoku puzzle [7]

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

Figure 2: The sudoku puzzle in figure 1, completed [7]

What is a Latin Square? A Latin square is an $n \times n$ square, filled with n different symbols, that appear in every row and column, exactly once. An example of a Latin square with $n = 3$ is shown in Figure 3:

1	2	3
2	3	1
3	1	2

Figure 3: A 3×3 Latin square

Latin squares can be traced back to Euler in the 18th century, but have earlier origins in the work of 17th century Korean mathematicians[4]. Earlier variations of Sudoku appeared in newspapers in France called *"carré magique diabolique"* [2]- meaning "Diabolical Magic Square", as seen in Figure 4. This was almost a modern Sudoku, but with some key differences. Whilst the 9×9 structure is similar, numbers could be repeated in rows and columns. Moreover, instead of the standard Sudoku rules, it followed the constraints of magic squares, requiring specific rows, columns, or diagonals to sum to a certain number.

Modern Sudoku puzzles, first appeared in puzzle books published by Dell Magazines - a company based in New York, under the name *Number Place*. This was thought to be the work of an American named Howard Garns in the late 1970's [15].

UN PROBLÈME PAR JOUR

5129. — CARRÉ MAGIQUE DE 9 A COMPARTIMENTS ÉGAUX

Composé par M. le commandant Coccoz

Compléter le carré ci-dessous en employant les 81 premiers nombres de manière que les neuf carrés composés de neuf nombres donnent chacun, comme total, 369. De plus, chacune des deux grandes diagonales, chaque horizontale et chaque verticale du carré complet devra donner 369.

17	20	5	[illegible]	67	79	54	36	39
70	73			42			8	11
45		33	2		26			64
19		16	41		29	63		78
47	32			81			10	22
75		72	13		1	35		50
59			27		15	37		34
6	18			28			77	62
31	43	46	80	56	68	12	24	9

Figure 4: A Diabolical magic square [2]

1.2 Where does Sudoku originate?

Sudoku as we know it now, was popularised by Maki Kaji of the Japanese games publisher Nikoli in 1984- it was here that the name *Sudoku* was conceived. Sudoku is an abbreviation for the phrase *sūji wa dokushin ni kagiru* [11], meaning "the numerals must remain single". Breaking this phrase down further: *sūji* means "number", and *dokushin* refers to "single status". Combined, these give us the term *Sudoku* (or *Single Numbers*), which has been its global name since the 2000s.

A New Zealander, Wayne Gould, stumbled upon the puzzle on a trip to Tokyo in 1997, and subsequently spent the next few years developing computer programs capable of generating Sudoku grids. It is significant that Sudoku grids can be filled with any 9 distinct symbols or numbers from scripts other than the standard Latin alphabet used widely in the Western World, making it a puzzle that transcends language barriers- further contributing to the puzzle's extensive popularity. Gould eventually brought the puzzle over to London.

Western newspapers, such as The Times popularised Sudoku as a regular newspaper puzzle in the early 2000s and today, it has become a staple of the newspaper puzzle scene, having exploded with popularity [15].

1.3 Why?

The maths behind Sudoku is significant because it allows us to:

- Analyse the puzzle structure
- Calculate the number of possible solutions and grids
- Determine the minimum number of clues needed for a unique solution
- Understand the symmetries within the grid

And much more by using concepts from combinatorics and group theory. Sudoku essentially provides a concrete, tangible example of these ideas in maths, that can be explored and understood through something generally accessible and comprehensible for most people.

1.4 What questions can we answer?

Mathematicians can analyse the puzzle's structure, calculate the number of possible solutions, determine the minimum number of clues needed for a unique solution, and understand the symmetries within the grid, all using concepts like combinatorics and group theory, which are fundamental mathematical fields with applications beyond just puzzles. Essentially, Sudoku provides a concrete example of complex mathematical ideas that can be explored and understood through a fun and accessible game for most people. We can answer many questions such as:

- How many Sudoku grids exist?
- How many of these are essentially different?
- What is it to be essentially different vs distinct?
- What ways can grids be symmetric?
- How many clues are needed to solve a Sudoku grid?

1.5 Some basic facts

Previously by Felgenhauer and Jarvis[6], it has been determined that there are

$$6670903752021072936960 \approx 6.67110^{21}$$

distinct Sudoku grids. However, calculating the grids that are essentially different is a different task. Prior to this, we should understand how *essentially different* differs from *distinct*. If you can produce a grid, from another grid by applying one or more transformations, they are equivalent, or distinct. If there is no such sequence of symmetries that map one grid to another, then the grids are *essentially different*. There are known to be 5472730538 *essentially different* Sudoku grids [14].

When counting if grids are essentially different we must consider if certain transformations, which we'll expand on in section 3, can preserve symmetries, whilst maintaining the integrity of a Sudoku grid. Filled grids can be transformed to other grids in a very simple manner. For example, if we take and grid and rotate it 90° clockwise, the grid itself is different, but it can be noted that the resulting grid is still a valid Sudoku solution.

2 Counting Shidoku

2.1 Shidoku

An easier way to count the number of grids is by using the 4×4 case of Sudoku- referred to most commonly as a *Shidoku* grid, where *shi* means "four" [16]. As seen in 1.5, the number of essentially different, let alone distinct grids is immense, thus is impossible to compute by hand. Although, for Shidoku grids this number is considerably smaller, making it far easier to compute manually. There are also less transformations and regions to consider. It can be shown that there are 288 distinct grids and only 2 essentially different Shidoku grids, which we will look at in section 2.2. The same concepts apply; each row and column must contain the numbers 1 to 4 exactly once, as do each 2×2 sub-grid. An example of a filled Shidoku grid is shown in figure 5.

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

(a) Valid Shidoku solution.

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

(b) Another valid solution.

Figure 5: Different valid Shidoku solutions.

A valid solution to a Shidoku grid is a 4×4 Latin square, with additional sub-grid constraints. Without considering the rules of Sudoku, each cell of a 4×4 grid can be filled 4^{16} ways with 4 unique numbers or symbols. Before requiring brute-force calculations, it is important to reduce the number of grids that may be equivalent.

2.2 Distinct Shidoku grids

Initially let us consider a non-optimal approach to counting the number of Shidoku grids. Any 4×4 grid could have

$$4!^4 = 24^4 = 331,776$$

ways to fill it. This is simply from filling in the rows, whilst ignoring the constraints of Sudoku. As seen this number is immensely large. We can refine this number further by considering the quantity of 4×4 Latin squares of which there are 576. As mentioned previously, any valid Sudoku solution is a Latin square. Now only the sub-grid constraints

1	2		
3	4		

Figure 6: Initial arrangement of the first sub-grid

are left to consider, which will result in 288 distinct grids. We can illustrate this step-by-step by filling in the grids manually.

To begin, we can fill the upper-left 2×2 sub-grid with the numbers $\{1, 2, 3, 4\}$ in $4!$ different ways, as shown in Figure 6. Next, we complete the first row and first column.

1	2	3	4
3	4		
2			
4			

Figure 7: There are 2 choices for the first column and row resulting in $4! \times 2 \times 2$ possibilities so far

At this stage, the grid can still be completed in multiple ways. Since each row and column must contain unique values, we find that each has precisely 2 ways to be completed. Thus, we have

$$4! \times 2 \times 2$$

ways to arrange the grid so far, shown in Figure 8

1	2	3	4
3	4		
2		4	
4			

Figure 8: Applying the rules of Sudoku to determine the placement of the final 4

By the basic rules of Sudoku (each row, column, and 2×2 sub-grid must contain the numbers $1 - 4$ without repeating any number), the 4 placed in the bottom-right sub-grid in Figure 8 is the unique solution to that cell. We can then logically complete the rest of the puzzle as seen in Figure 9. Since only the numbers $\{1, 2, 3\}$ remain to be placed, we find that there are exactly 3 ways to complete the grid. Therefore, the total number of distinct Shidoku grids is:

$$4! \times 2 \times 2 \times 3 = 288$$

i.e., there are 288 different ways to fill in a Shidoku puzzle

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

(a) Solution 1

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

(b) Solution 2

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

(c) Solution 3

Figure 9: The three final ways to fill the grid.

However it is important to note, that the specific numbers placed in the grid are arbitary, rather it is the relative positions of differently numbered cells to one another that define each distinct grid configuration. Regardless of how the cells are filled in initially, we will reach a stage where only 3 grids (that are fundamentally different) remain. Although 288 distinct Shidoku grids exist, many of these grid configurations are structurally the same, or invariant, under various transformations, such as relabelling, rotations or reflections. By considering only the structure, we can provisionally conclude that there are 3 essentially different Shidoku grids. This will be explored further in section 3.3.

3 Transformations

3.1 Which transformations can be applied?

A Shidoku grid remains structurally identical under certain transformations, which form a symmetry group. These transformations that preserve the rules of Sudoku include:

- Swapping columns **within a sub-grid**
- Swapping rows **within a sub-grid**
- Rotations (0°, 90°, 180°, 270°)
- Reflections (horizontal, vertical, diagonal)
- Swapping entire row groups
- Swapping entire column groups

Along with relabelling, which refers to replacing one set of symbols with another, these form a group. Each of these symmetries permute the set of 288 Shidoku grids to one another, whilst preserving the sudoku constraints (i.e the resulting grid is a valid sudoku or Shidoku solution).

We can define a **group** in the following way:

Definition 3.1 (Group)**.** A **group** $(G, *)$ is a non-empty set G with a binary operation $*$ that satisfies the following properties:

- **Associativity:** For all $a, b, c \in G$,

$$(a * b) * c = a * (b * c).$$

- **Identity element:** There is an element $e \in G$ such that for all $g \in G$,

$$g * e = e * g = g.$$

- **Inverses:** For every $g \in G$, there is an element $g^{-1} \in G$ such that

$$g^{-1} * g = g * g^{-1} = e.$$

The operation $*$ is called the *group operation*, and g^{-1} is the *inverse* of g.

Binary operations are simply those such as $\times$ or $+$ that act on two numbers to produce another.

3.2 Effects of Transformations

Whilst the operations outlined in section 3.1, will always generate valid Shidoku grids, not all transformations do this. For example take a standard Shidoku grid:

- Swapping the 2nd and 3rd or 4th rows or columns arbitrarily does not necessarily maintain validity.
- Swapping the 1st and 3rd or 4th rows or columns does not either.
- However, swapping the top two rows as a unit with the bottom two preserves the structure.
- As does swapping the left two columns with the right two.

Previously in section 2.2, we said that the numbers/symbols/colours used to fill each cell are arbitrary, as the puzzle itself has not really changed. We can use relabelling as a good visualisation. Relabelling is simply swapping the symbols in each square with another. This produces a different grid labelling, but the same grid configuration. In Figure 10, every occurrence of 1 in the first square has been replaced by 2 in the second square. Every occurrence of 2 has been replaced by 3, and so on with every occurrence of 9 being replaced by 1 [13]. We can see here that the puzzle itself has not changed. From this we can answer further how many essentially different Shidoku grids exist.

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

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

Figure 10: A relabelled Sudoku grid [13]

3.3 Essentially Different Shidoku grids

Now that we have a better understanding of the transformations that can be applied to Shidoku and Sudoku grids, we can go back to Figure 9 in section 2.2, and can look at solutions 2 and 3 to determine the *essentially different* grids

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

(a) Solution 2

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

(b) Solution 3

Figure 11: Solutions from figure 9

Here we can relabel solution 2 so that every occurrence of 2 is replaced by 3, and every occurrence of 3 is replaced by 2. This results in Figure 12 below:

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

(a) Solution 2

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

(b) Solution 3

Figure 12: Solutions from Figure 9, where in Solution 2, all occurrences of 2 and 3 have been swapped. Solution 3 remains unchanged.

Now, we can take solution 2 and apply a reflection in the main diagonal (upper-left to bottom-right), seen in Figure 13.

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

(a) Solution 2

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

(b) Solution 3

Figure 13: Solutions from Figure 9, where Solution 2 has been reflected along the main diagonal (top-left to bottom-right). Solution 3 remains unchanged.

By applying this combination of transformations, we can see that solutions 2 and 3 are actually distinct, not essentially different. Therefore we can state that out of 288 distinct Shidoku grids, only 2 are essentially different. Additionally we can say that all 288 distinct Shidoku grids have been mapped from either solution 1 or 2. The symmetry group discussed previously, act on the set of 288 grids and partition them into 2 orbits. We call these:

- **Type 1** grids, of which there are 96,
- **Type 2** grids, of which there are 192

These are illustrated in Figure 14.

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

(a) Type 1

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

(b) Type 2

Figure 14: Type 1 and Type 2 Shidoku grids

We can now consider subgroups of the symmetry group that acts on these orbits, and apply Burnside's Lemma to calculate the number of essentially different grids in an alternative manner.

4 Counting grids

In the previous section, we manually calculated the number of essentially different grids by filling in Shidoku grids. However group theory can provide a more refined and systematic approach. By taking symmetry groups and determining their orbits, we can then apply Burnside's Lemma.

Burnside's Lemma, also referred to as Burnside's counting Theorem or Cauchy-Frobenius Lemma, is a useful result in Group Theory, that takes into account symmetry when counting objects in Maths. It was discovered by Cauchy and Frobenius, but was first referenced in William Burnside's 1897 book *"Theory of groups of finite order"* [3], from which thereafter it became more well-known.

Previous works, such as those by Arnold et al.[1] or by Russell and Jarvis [14], have used GAP to evaluate which grids remain fixed under transformations. By also taking the conjugacy classes of the symmetry groups, we can simplify these calculations provide a more efficient count, which we will explore in this section.

4.1 Burnside's Lemma

To understand the application of Burnside's lemma there are a few concepts that we will introduce, in the context of Shidoku.

Definition 4.1 (Symmetry Group)**.** A **symmetry group** is a group in which its elements are the transformations that can be applied to objects, and whose operation is the composition operation $\circ$.

A symmetry group can act on a set of Shidoku grids, transforming them, whilst preserving their structure.

Definition 4.2 (Orbit)**.** Let G be a group acting on a set X. Then for every $x \in X$, the set

$$\mathrm{Orb}_G(x) = \{g(x) \mid g \in G\}$$

is called the **orbit** of x. This represents all possible transformations of $x \in X$ under the group action.

Definition 4.3 (Stabiliser)**.** The **stabiliser** of x in G is given by

$$\mathrm{Stab}_G(x) = \{g \in G \mid g(x) = x\}$$

This consists of all elements of G that leave x unchanged.

Definition 4.4 (Fixed Point)**.** Let G be a group acting on a set X. Then an element $x \in X$ is called a **fixed point** for this action if $\mathrm{Orb}_G(x) = \{x\}$, or equivalently if $\mathrm{Stab}_G(x) = G$. We write

$$\mathrm{Fix}_G(X) = \{x \in X \mid \mathrm{Orb}_G(x) = \{x\}\}$$

for the set of fixed points.

In the context of Sudoku this refers to the Shidoku grids that remain unchanged under all transformations of G. These are all the definitions needed to understand Burnside's Lemma.

Lemma 4.5 (Burnside's Lemma)**.** *Let G be a finite group that acts on the set X. Let X/G be the set of orbits on X (that is, each element of X/G is an orbit of X). For any element $g \in G$, let X^g be the set of points of X that are fixed by g:*

$$X^g = \{x \in X \mid g \cdot x = x\}.$$

Then

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |X^g|$$

Burnside's Lemma simply states what the number of **essentially different** objects under a symmetry group G is. Where:

- $|G|$ is the number of symmetries (rotations, reflections, swaps e.t.c).
- X^g is the set of Shidoku grids unchanged by said symmetries $g \in G$.
- The sum counts the number of grids fixed by all group elements.
- $|X/G|$ is the number of essentially different grids. In other words, the number of orbits is the average number of fixed points of G.

We can use Burnside's, and find that the number of **essentially different** Shidoku grids is 2. To now understand the proof of this Lemma we will need the Orbit-Stabiliser Theorem.

Theorem 4.6 (Orbit-Stabiliser Theorem)**.** *Suppose G is a group acting on the set of Shidoku grids X, and let $x \in X$ be a specific Shidoku grid. Then the size of the orbit of x under the action of G is given by*

$$|\operatorname{Orb}_G(x)| = [G : \operatorname{Stab}_G(x)],$$

where $\operatorname{Stab}_G(x)$ is the stabiliser of x in G, i.e., the group of symmetries that leave the grid x unchanged.

In particular, if G is finite, then by Lagrange's Theorem, we have

$$|G| = |\operatorname{Orb}_G(x)| \cdot |\operatorname{Stab}_G(x)|.$$

This result is relevant in understanding the symmetries of Shidoku grids and can be used in the proof of Burnside's Lemma to count the number of essentially different grids under the action of a symmetry group with respect to composition.

Proof of Burnside's Lemma. We can start with the sum

$$\frac{1}{|G|}\sum_{g\in G}|X^g| = \frac{1}{|G|}\sum_{g\in G}|\{x \in X : gx = x\}|$$

By definition, X^g is rewritten in set notation.

$$= \frac{1}{|G|}\sum_{x\in X}|\{g \in G : gx = x\}|$$

Now we have changed the indexing, instead summing elements $x \in X$, and counting how many transformations g fix x. Then, by definition, $\mathrm{Stab}(x)$ is the stabilizer of x.

$$= \frac{1}{|G|}\sum_{x\in X}|\mathrm{Stab}(x)|$$

By the Orbit-Stabilizer Theorem (4.6), we replace $|\mathrm{Stab}(x)|$ with $\frac{|G|}{|\mathrm{Orb}(x)|}$. The orbit-stabilizer theorem gives: $|G| = |\operatorname{Orb}_G(x)| \cdot |\operatorname{Stab}_G(x)|$, so the sum becomes:

$$\begin{aligned}\sum_{x\in X}|\mathrm{Stab}(x)| &= \sum_{x\in X}\frac{|G|}{|\mathrm{Orb}(x)|}\\ &= |G|\sum_{x\in X}\frac{1}{|\mathrm{Orb}(x)|}.\end{aligned}$$

Finally, we get that:

$$\sum_{g\in G}|X^g| = |G|\cdot|X/G|.$$

and the lemma follows. □

4.2 Symmetry Groups

As we have mentioned previously 4.1, the symmetries of a Sudoku square can form a group under composition. We can take this symmetry group and apply Burnside's Lemma to count how many Shidoku or Sudoku grids are preserved by each of these transformations in a given symmetry group. The full set of Sudoku symmetries is quite large- specifically, it has order 3072 [1]. This is also the largest complete Shidoku symmetry group. A Shidoku symmetry group is *complete* if it's actions partition the set of 288 distinct Shidoku groups into two separate orbits, which we labelled earlier as **Type 1** and **Type 2** grids.

This full group of Shidoku symmetries can be useful, but it's size means that even with the help of a computer, the calculations are quite demanding. Instead we can consider a smaller subgroup of symmetries that can simplify the calculations needed to use Burnside's Lemma. These groups are still *complete*, as the subgroups still partition the set of distinct Shidoku grids into the valid orbits and we call them the "*minimal complete symmetry groups*" [1], as they are the smallest groups that sustain this property. We can now explore various symmetry groups, comparing their sizes. A well-structured symmetry group of smaller order, can make calculations streamlined whilst capturing the same structure.

To begin with, let us consider the group of row permutation that can be applied to a Shidoku grid, which we can define as a set of cycle permutations. The group H is generated by

$$H = \{(e), (1\ 2), (3\ 4), (1\ 3)(2\ 4)\},$$

where (1 2) refers to swapping the first and second rows, (3 4) swaps the third and fourth rows, and (1 3)(2 4) swaps the two row blocks. These elements can be applied to find the full H group, which is

$$\{e, (1\ 2), (3\ 4), (1\ 3)(2\ 4), (1\ 2)(3\ 4), (3\ 4)(1\ 2), (1\ 4\ 2\ 3), (1\ 3\ 2\ 4)\}.$$

The element (e) represents the identity, i.e., doing nothing. The full group H has order 8 and is isomorphic to the dihedral group D_4, which describes the symmetries of a square. Why is it relevant that H is isomorphic to D_4? The relevance of H being isomorphic to D_4 stems from the fact that D_4 consists of 8 elements, which include both rotations and reflections of a square. Specifically, these transformations are:

$$\{e, r, r^2, r^3, s, sr, sr^2, sr^3\},$$

where r represents a 90° rotation clockwise and s represents a reflection (e.g., across the horizontal, vertical or diagonal axes). These symmetries represent the full set of rotations and reflections that preserve the structure of a square.

Since H is isomorphic to D_4, this suggests that the operation of H on the set of Shidoku grids can be described in a similar manner to the actions of D_4. Essentially, by understanding the structure of D_4, we can easily comprehend how the symmetries of H act on the Shidoku grids. This simplifies the process of analysing and counting the number of distinct grids preserved by these symmetries using Burnside's Lemma.

In group theory, D_4 can be written as

$$D_4 = \langle r, s \mid r^4 = e, s^2 = e, srs = r^{-1} \rangle,$$

where r and s are the generators of the group, and for the relations of the rotations and reflections within the dihedral group. This structure allows for a simpler approach to counting Shidoku grids rather than using the full set of Shidoku symmetries.

To show that H is isomorphic to D_4, we can define a mapping $\varphi : D_4 \to H$ as shown:

$$\varphi(r) = (1\ 4\ 2\ 3), \quad \varphi(s) = (3\ 4),$$

where $r = (1\ 4\ 2\ 3)$ is the 90° rotation and $s = (3\ 4)$ is the reflection. We first check that this mapping is a well-defined homomorphism, meaning that it follows the group operation. For any elements $d_1, d_2 \in D_4$, we need to confirm that:

$$\varphi(d_1 d_2) = \varphi(d_1)\varphi(d_2).$$

by definition of an isomorphism. Also the known relations of D_4, such as $r^4 = e$, $s^2 = e$, and $srs = r^{-1}$, must be verified to confirm that the mapping preserves the group structure:

1. $\varphi(r^4) = \varphi(e) = e_H$ (identity in H), since $r^4 = e$ in D_4 and the identity in H is e_H.
2. $\varphi(s^2) = \varphi(e) = e_H$ (identity in H), because $s^2 = e$ in D_4, and we map it to the identity in H.
3. $\varphi(srs) = \varphi(r^{-1})$, which corresponds with elements in H. In D_4, $srs = r^{-1}$, and in H, $(3\ 4)(1\ 4\ 2\ 3)(3\ 4) = (1\ 3\ 2\ 4\ )$, which corresponds to the inverse operation of $(1\ 4\ 2\ 3)$.

We have therefore shown that distinct elements in D_4 have a mapping to a distinct element in the co-domain H. Thus, φ is a well-defined map and so a homomorphism.

Next, we show that φ is surjective, which means that every element of H has a corresponding element in D_4. By mapping the generators r and s, we can generate all elements of D_4, and we can also find corresponding elements in H. Therefore, all elements of H can be mapped from elements in D_4, confirming that φ is surjective. Since φ is a homomorphism and $|D_4| = |H|$ (both are of order 8), the kernel of φ must be trivial (i.e., the only element in D_4 that maps to the identity in H is the identity element e). Hence, φ is injective. Since φ is both surjective and injective, it is thus bijective. Being a bijective homomorphism, φ is then said to be an isomorphism between D_4 and H. Thus, H is isomorphic to D_4, and the group H preserves the same structure as D_4, allowing us to use the isomorphism to analyse the symmetries of Shidoku grids. One method we can use, is through the conjugacy classes of H. As $D_4 \cong H$, we can also map the conjugacy classes of H from D_4, as they will be equivalent.

4.3 Conjugacy classes

Definition 4.7 (Conjugacy Class)**.** A **conjugacy class** of a group, is a set of elements that are connected by conjugation. Let G be a group and let $x \in G$. Then the conjugacy class of x, denoted x^G, is defined by

$$x^G = \{gxg^{-1} : g \in G\}.$$

This operation is defined in the following way: Two elements $x, y \in G$ are said to be conjugate in G if there exists $g \in G$ such that $y = gxg^{-1}$.

Why can we use conjugacy classes? In a group, the position symmetries within the same conjugacy class, will preserve the same number of Shidoku or Sudoku boards. This massively simplifies the calculations needed for Burnside's Lemma. Russell and Jarvis [14] used a computer to apply Burnside's Lemma to the full complete group of position symmetries.

Their brute-force computation found 5,472,730,538 equivalent sudoku grids, through 275 conjugacy classes from the symmetry group of order 3359232 [14]. So instead of checking which grids are preserved from the set of 3359232 symmetries, which would require an immense amount of time, they could select one from each conjugacy class. Therefore, it was only necessary to check what would occur for those 275 symmetries. However, as mentioned previously, taking a smaller, but complete, symmetry group can simplify these calculations. These are referred to as *minimal complete symmetry groups* [1].

For example take H from earlier. We can take the conjugacy classes of D_4, of which there are known to be 5, and map the conjugacy classes of H from there. The conjugacy classes of D_4 are

$$\{e\}, \{r, r^3\}, \{sr, sr^3\}, \{r^2\}, \{s, sr^2\}$$

From this, we can determine the conjugacy classes of H, which are

$$\{e\}, \{(1\ 3)(2\ 4)\}, \{(1\ 4\ 2\ 3)\}, \{(1\ 3\ 2\ 4)\}, \{(1\ 2), (3\ 4)\}, \{(1\ 2)(3\ 4), (3\ 4)(1\ 2)\}$$

This reduces the symmetries that need to be applied to Burnside's from 8 to 5. From this class of row swaps; $\{(1\ 2), (3\ 4)\}$, only one would need to be applied. Another good example of a minimal complete symmetry group comes from Arnold et al. [1].

A *minimal complete Shidoku symmetry group* is formally a subgroup of the full symmetry group, that preserves the structure of the two orbits of Shidoku boards while simplifying the structure of the group. A full Shidoku symmetry group G_4, from $G_4 = H_4 \times S_4$, has order 3072. It is the direct product of H_4, the group of position symmetries, and S_4, the group of relabelling symmetries, which acts on the set of 288 Shidoku boards. Here the elements of S_4 are those which permute the values $\{1, 2, 3, 4\}$ on the board, and are called the *relabelling symmetries*, and elements of S_{16} that permute the cells of the board while maintaining the Shidoku conditions of the grids, are called *position symmetries*. These are further discussed in Arnold et al. [1]. Think of each Shidoku cell being labelled 1 to 16, as seen in Figure 15.

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

Figure 15: Labelled Shidoku grid

Then the transpositions t of S_{16} can be written in cycle notation here:

$$(2\ 5)(3\ 9)(4\ 13)(7\ 10)(8\ 14)(12\ 15)$$

The position symmetry group H_4 is generated by: swapping rows/columns between bands/pillars, swapping bands/pillars, rotations r of the board by a quarter turn clockwise and transpositions t of the board.

If we let s to be the swap of the third and fourth rows of a Shidoku board, then H_4 is generated by r, s, and t. The full presentation of the position symmetry group is

$$H_4 = \langle r, s, t \mid r^4, s^2, t^2, trtr, rsr^2tsr^3t, tststs, srsr^3srsr^3 \rangle.$$

We will use the more compact notation $H_4 = \langle r, s, t \rangle$. This is a non-abelian group of order 128 , so the full Shidoku symmetry group $G_4 = H_4 \times S_4$ has order

$$128 \times 4! = 3072.$$

The minimal complete symmetry group $\langle s, t \rangle \times S_4$ from Arnold et al.[1], which is a subgroup of H_4, can be written as

$$\langle s, t \mid s^2, t^2, (st)^4 \rangle.$$

where (s, t) corresponds to the position symmetries that act on the board. This is then presented as

$$\{s, t\} = \{\ e, s, t, st, ts, sts, tst, stst\} \subseteq H_4$$

which is generated by $\langle s, t \mid s^2, t^2, (st)^4 \rangle$. This group is also referred to as a *Coxeter group.*

Definition 4.8 (Coxeter group)**.** A **Coxeter group** is an abstract group with the presentation,

$$\langle r_1, r_2, \ldots, r_n \mid (r_i r_j)^{m_{ij}} = 1 \rangle$$

where $m_{ii} = 1$ and $m_{ij} = m_{ji} \geq 2$ is either an integer or ∞ for $i \neq j$.

These groups, named after H.S.M Coxeter and introduced in 1934, [5] are simply those that can be described by a set of generators (usually basic reflections) and relations. D_4 is also a Coxeter group.

The group $\langle s, t\rangle \times S_4$, similar to H, is also isomorphic to D_4, and so has 8 elements and 5 conjugacy classes. This mapping can be shown similarly to the well-defined map $\varphi : D_4 \to H$, where elements $d \in D_4$ have corresponding elements in $\langle s, t\rangle \times S_4$. Where H is a subgroup of the full complete symmetry group $\langle H \times H\rangle \rtimes C_2 \times S_4)$, and $(s, t) \times S_4$ is a subgroup of a different full complete symmetry group $G_4 = H_4 \times S_4$. The full complete symmetry group $(H \times H) \rtimes C_2 \times S_4)$ also has order 3072. $|H| = 8$, so $|\langle H \times H\rangle \rtimes C_2 \times S_4)| = 8 \times 8 \times 2 \times 4! = 3072$.

This reduction in group size not only makes the analysis easier but still provides the full symmetry structure of the two orbits. We can now, using results from Arnold et al. [1], complete the Burnside's calculations. We can partition the subgroup $\langle s, t\rangle \times S_4$, into five conjugacy classes, which are

$$\{e\}, \{s, tst\}, \{t, sts\}, \{st, ts\}, \{stst\}$$

From each conjugacy class, we then take the representative elements : $(e, s, t, st, stst)$. Each of these elements will preserve the same number of Shidoku grids, as the others in each conjugacy class. To clarify, s refers to the swap of the third and fourth rows, and t is essentially the transposition of the board, where the grid is reflected in the main diagonal from the upper-left to bottom-right.

For ease of visualisation, we can represent these symmetries in cycle notation. First, let us refer to the grid from Figure 15 again, where each cell in a Shidoku grid has been labelled with the numbers $1 - 16$. It is shown again here in Figure 16.

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

Figure 16: The labelled Shidoku grid from Figure 15

The symmetry s can be represented as the cycle,

$$(9\ 13)(10\ 14)(11\ 15)(12\ 16)$$

the reflection t as,

$$(2\ 5)(3\ 9)(4\ 13)(7\ 10)(8\ 14)(12\ 15)$$

the symmetry st as,

$$(2\ 5)(3\ 9\ 4\ 13)(7\ 19\ 8\ 4)(11\ 12\ 16\ 15)$$

and $stst$ as,

$$(3\ 4)(7\ 8)(9\ 13)(10\ 14)(11\ 16)(12\ 15)$$

.

Now using *MATLAB* and *GAP*, the fixed grids can be found. This is shown in Figure 17 below.

Conjugacy class	**Representative element**	**Invariant grids**
$\{e\}$	e	$12 \cdot 4! = 288$
$\{s, tst\}$	s	$0 \cdot 4!$
$\{t, sts\}$	t	$2 \cdot 4! = 48$
$\{st, ts\}$	st	$0 \cdot 4!$
$\{stst\}$	$stst$	$0 \cdot 4!$

Figure 17: Conjugacy classes and the Shidoku grids they fix [1]

As expected the identity e, preserves all 288 Shidoku grids. Then no other element except, t preserves any symmetries. From applying Burnside's Lemma (4.5)

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |X^g|$$

to the data in Figure 17, we can see that the number of essentially different Shidoku grids is indeed 2. So we can say that $\langle s, t \rangle \times S_4$, is indeed a minimal complete symmetry group, as it has been shown to produce 2 known orbits of the set of distinct Shidoku grids, that we know as **Type 1** and **Type 2**. This calculation is

$$\frac{1(12 \cdot 4!) + 2(0) + 2(2 \cdot 4!) + 2(0) + 1(0)}{8 \cdot 4!} = 2$$

Where $8 \cdot 4!$ is the order of $\langle s, t \rangle \times S_4$.

4.4 Graphical representation

This result can also be represented graphically, and has shown to be using the graph visualisation program *yEd*, by Arnold et al. [1].

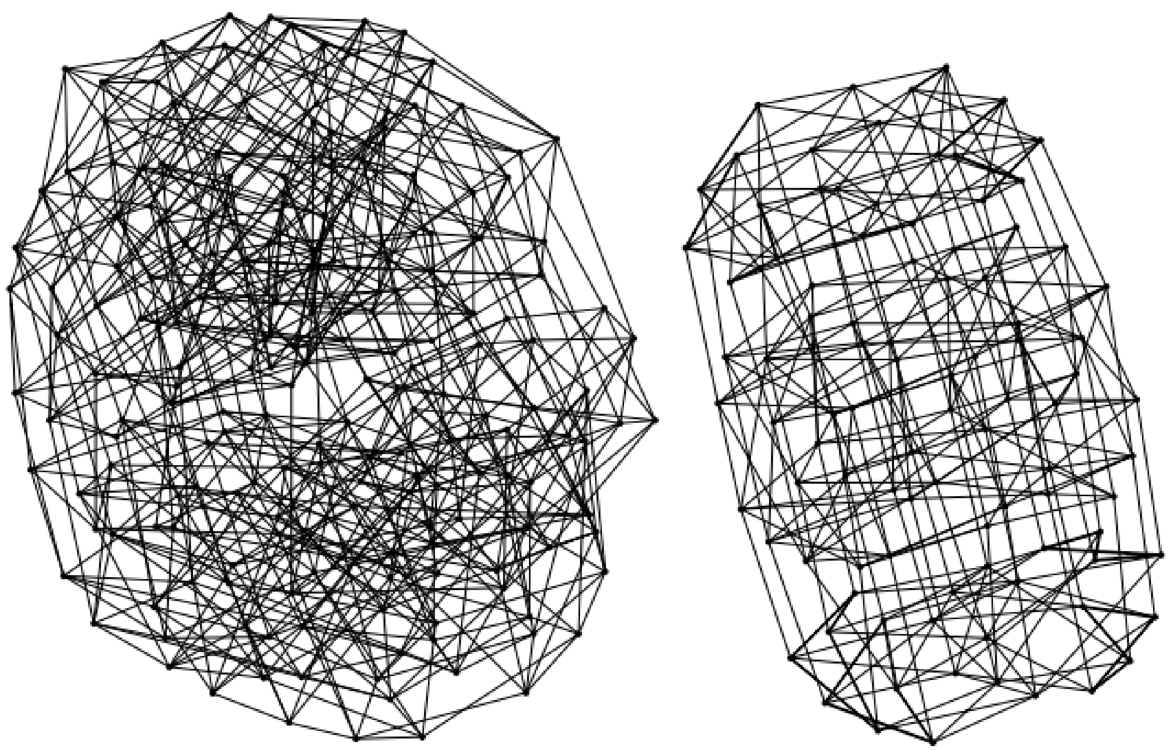

Figure 18: The full Shidoku symmetry group $G_4 = H_4 \times S_4$

Here, Figure 18 shows the full Shidoku symmetry group $G_4 = H_4 \times S_4$ acting on the set of 288 distinct Shidoku grids. Each node, represents an individual grid, and every edge illustrates all transformations that map one grid to another. This graph shows quite a high degree of connectivity. It is clearly shown here that the symmetry group $G_4 = H_4 \times S_4$, partitions the set of distinct Shidoku grids into the two orbits previously mentioned. The graph on the left, which is larger and more connected, represents the larger set of **Type 2** grids of which there are 192, and the graph on the right corresponds to the smaller set of **Type 1** grids of which there 96.

Now for the minimal complete Shidoku symmetry group $\langle s, t\rangle \times S_4$, shown in Figure 19, where we can see that this group partitions the set of essentially different Shidoku grids into two distinct orbits. A key difference between the graphs in Figure 18 and Figure 19 is that Figure 19 is less highly connected. This is because we have used a smaller symmetry group, so whilst all 288 distinct Shidoku grids are represented by a node, not all the mappings between them are graphed. However not all subgroups of $G_4 = H_4 \times S_4$, are minimal complete symmetry groups.

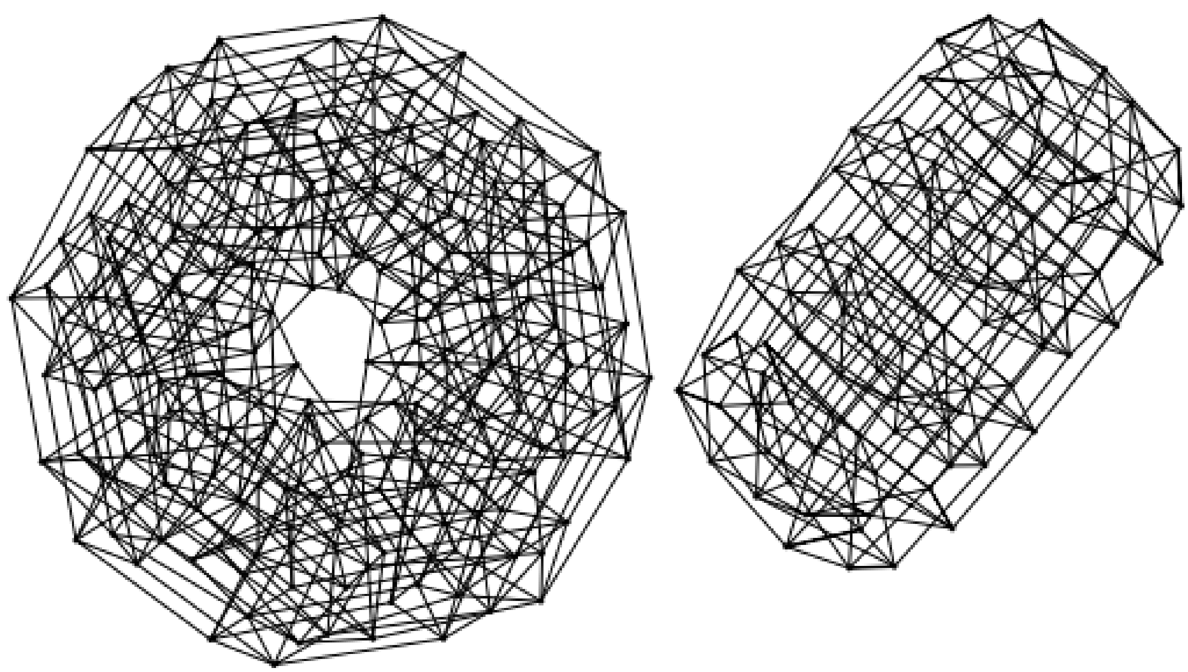

Figure 19: The minimal complete Shidoku symmetry group $\langle s, t\rangle \times S_4$

Some groups of larger size than $\langle s, t\rangle \times S_4$, partition the set of 288 Shidoku grids into more than 2 orbits. However, some which are *minimal*, are not necessarily also *complete*. For example, take $\langle r, t\rangle \times S_4$, which is the subgroup of order 8, represented by $\langle r, t | r^4, t^2, trtr\rangle \subseteq H_4$, which is also explored further in Arnold et al. [1]. Although $|\langle s, t\rangle \times S_4| = |\langle r, t\rangle \times S_4|$,

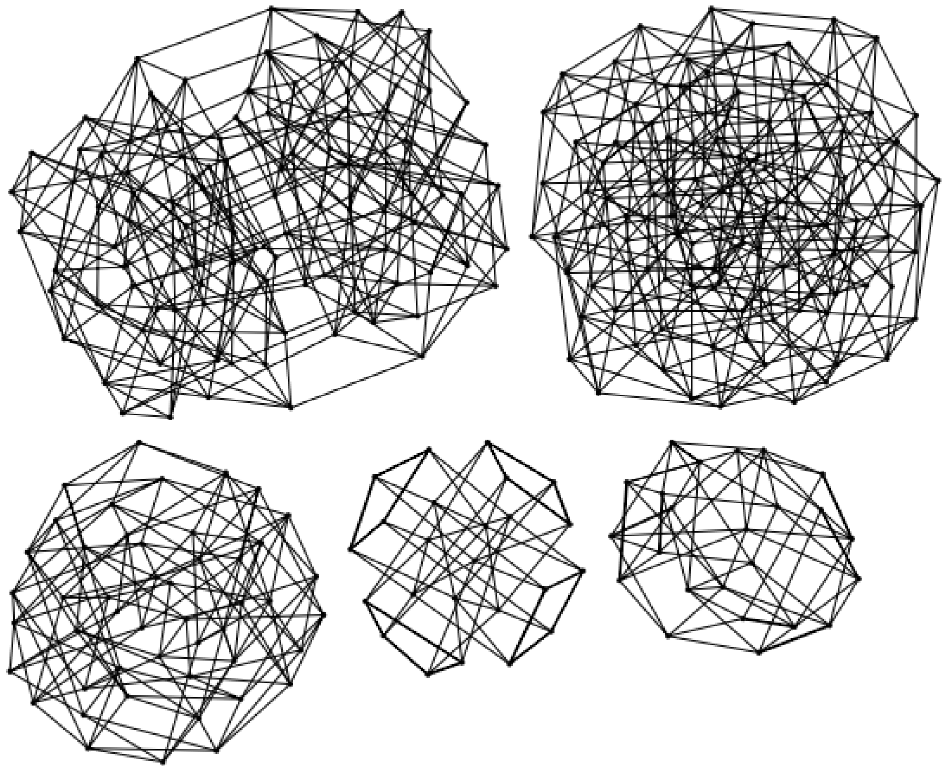

Figure 20: The action of the subgroup $\langle r, t\rangle \times S_4$

the action of this group results in 5 orbits which are represented graphically in Figure 20. Therefore, we cannot describe this symmetry as a minimal complete symmetry group. Our motivation for a symmetry group being minimal, is simply efficiency. A smaller, well-chosen symmetry group simplifies Burnside's lemma calculations, making them easier to interpret and far less computationally demanding. This reasoning can be extended to the 9×9 Sudoku case. If instead of the full group of 3359232 symmetries, and 275 resulting conjugacy classes enumerated by Russell and Jarvis [14], a smaller complete subgroup would allow for a clearer and more structured calculation. This idea is further explored in Arnold et al. [1].

5 Solving Puzzles

Thus far, we have examined the structure of Sudoku grids of varying sizes and how they are formed. We can also explore how to approach solving them. A Sudoku puzzle of any size generally already has some cells pre-filled; otherwise the puzzle would present itself as an empty grid that could be filled in 6.67110^{21} different ways. When looking at the number of clues, or *givens*, in a Sudoku board, we can answer some compelling questions:

- What is the minimum number of clues/givens needed for a Sudoku puzzle to have a unique solution and be solvable?
- How can a puzzle be solved using algorithms?
- What are some logical ways to solve a puzzle by hand?

However, we should understand first that a puzzle must have a sufficient quantity of givens, in order for the ensuing solution to be proper or **well formed**.

5.1 Counting Clues

A **well formed** Sudoku puzzle refers to one that has a unique solution. If there are too few clues provided in a puzzle, then multiple solutions will follow, and the resulting Sudoku board will not be proper. On average, most newspapers and magazines publish 9×9 Sudoku puzzles with 25 clues [8]. The Japanese games publisher Nikoli, mentioned in section 1.2, has a policy that no puzzle published by them will contain more than 32 clues.

This problem was studied extensively in 2014 by McGuire, Tugemann and Civario [8], where they validated that the minimum number of clues required for a Sudoku puzzle to have a unique solution, is 17. Whilst it was widely accepted that 17 was a suitable lower bound for this problem, proving that there was no such valid 16-clue grid was considered a monumental task, that many had previously attempted.

McGuire, Tugemann and Civario proved this with software they developed named *checker* in `C++`. Initially the program was quite slow, and would have taken approximately 300,000 processor years to solve. They eventually optimised this algorithm that meant, *checker*, would take approximately 800 processor years instead [8]. Their approach involved:

1. Creating a catalogue of all completed Sudoku grids (with some restrictions)
2. Developing a checker program by using efficient algorithms to search for extensively, a completed Sudoku grid, for a puzzle containing 16 clues and confirming that it had a unique solution
3. Using this checker to search through the entire catalogue of completed Sudoku puzzles.[8]

Although, a 17-clue puzzle will always ensure a unique solution, the converse does not always hold; a puzzle with a unique solution does not guarantee that it originated from a grid with 17 clues. Different Sudoku configurations will have varying initial clue placements. Some grids, if they contain less than a certain number of givens, will not yield a unique or proper solution and so is considered **irreducible** if in such a form.

For instance, take the example in Figure 21, from *Taking Sudoku Seriously* [13]. We can look at a grid that initially contains 26 clues, with the ones highlighted in green being removable either individually or in pairs of 1 and 7 or 1 and 9, reducing to a minimum of 24 clues. However, removing any more, would result in a puzzle with a non unique solution, making the puzzle **weakly completable**. In this form with 24 clues, shown in Figure 22 the puzzle is irreducible.

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

Figure 21: 26-clue Sudoku puzzle [13]

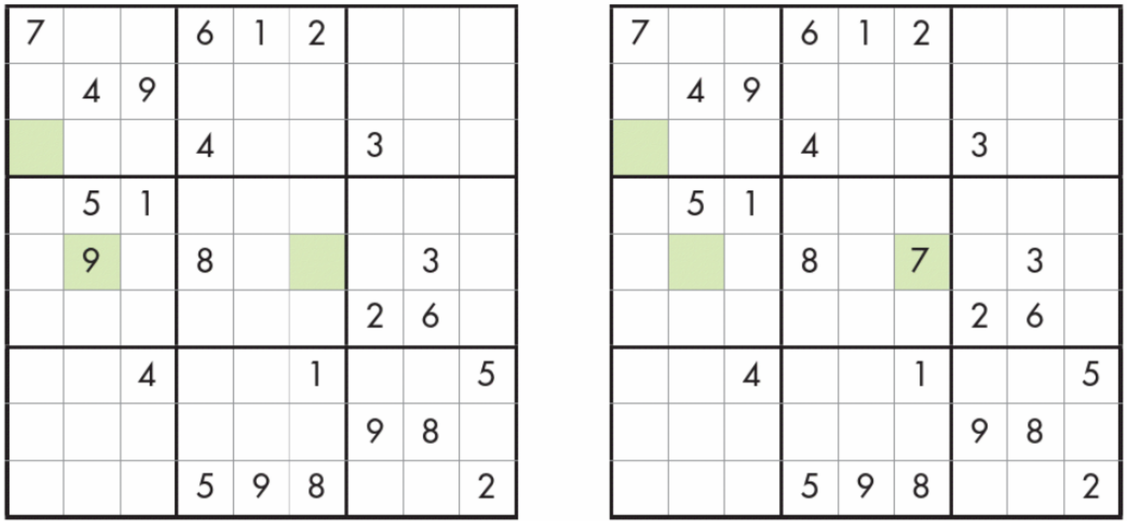

Figure 22: 24-clue Sudoku puzzles [13]

5.2 Solving Strategies

As mentioned previously, if a puzzle does not have a unique solution it is considered **weakly completable**. But what does this mean? A Sudoku puzzle can be either **strongly** or **weakly completable**. Adopting the terminology in [9], we can say a Sudoku grid is:

- **Strongly completable** if it can be solved alone through logical deduction, without any guessing.
- **Weakly completable** if some degree of search or trial-and-error is required to reach a solution.

5.2.1 Backtracking Algorithms

One method to solve puzzles is through backtracking algorithms [12]. Players can apply a broad range of computer algorithms to rapidly solve Sudokus. However for an $n \times n$ puzzle, as n increases in size, **combinatorial explosion** occurs. **Combinatorial explosion** refers to the rapid growth of the number of solutions due to the increasing complexity of the constraints and bounds of Sudoku as n increases. This significantly limits the size and number of grids that can be solved and analysed through brute-force enumerations.

In Backtracking, a brute-force algorithm will visit the empty cells of a Sudoku grid in some order, filling in numbers in order. If for example a cell is filled with 1, and this is not valid, the algorithm will place 2 and check again. If all numbers $1 - 9$ are cycled through, and none are valid, the algorithm will backtrack to the previous cell, and the same process repeats until a valid solution is found.[12]

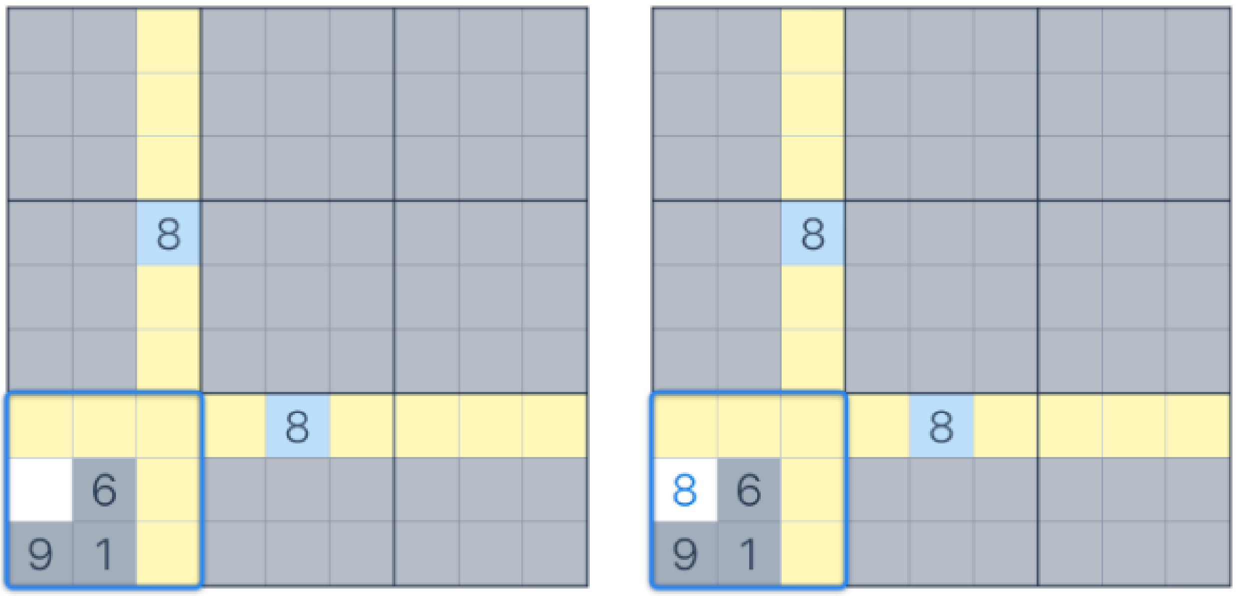

Figure 23: An example of the blocking effect [10]

5.2.2 The Blocking Effect

There also exists logical approaches in solving Sudoku puzzles. The *blocking effect* is one that involves searching a grid for situations where a number exists already in a row, column or block and then using this information to deduce that said number can only be placed in the remaining cells. This is essentially just confirming where numbers CANNOT be placed, and eliminating possibilities for those cells. This can be seen in the example in Figure 23.

We can see that the number 8 is placed in the third column and seventh row. This means that cells in which the remaining 8s can be placed in are limited. The cells highlighted in yellow show where an 8, cannot be placed. By process of elimination, this leaves only one square, in the bottom-left sub-grid, where an 8 can be placed. All puzzles, that are well formed can be solved logically, in this manner, as well as through a variety of modern strategies that have been developed over the years.

References

[1] Elizabeth Arnold et al. "Minimal complete Shidoku symmetry groups". In: *arXiv preprint arXiv:1302.5949* (2013).

[2] Christian Boyer. *Sudoku's French Ancestors*. 2006. URL: `http://www.multimagie.com/indexengl.htm` (visited on 01/03/2025).

[3] William Burnside. "Theory of groups of finite order". In: *Messenger of Mathematics* 23 (1909), p. 112.

[4] Charles J Colbourn. *CRC handbook of combinatorial designs*. CRC press, 2010.

[5] H. S. M. Coxeter. "Discrete Groups Generated by Reflections". In: *Annals of Mathematics* 35.3 (1934), pp. 588–621. ISSN: 0003486X, 19398980. URL: `http://www.jstor.org/stable/1968753` (visited on 03/11/2025).

[6] Bertram Felgenhauer and Frazer Jarvis. "Mathematics of sudoku I". In: *Mathematical Spectrum* 39.1 (2006), pp. 15–22.

[7] Sandiway Fong. *Sudoku Sandiway*. URL: `https://sandiway.arizona.edu/sudoku/examples.html` (visited on 01/03/2025).

[8] Bastian Tugemann Gary McGuire and Gilles Civario. "There Is No 16-Clue Sudoku: Solving the Sudoku Minimum Number of Clues Problem via Hitting Set Enumeration". In: *Experimental Mathematics* 23.2 (2014), pp. 190–217. DOI: `10.1080/10586458.2013.870056`. eprint: `https://doi.org/10.1080/10586458.2013.870056`. URL: `https://doi.org/10.1080/10586458.2013.870056`.

[9] A. D. Keedwell. "Two Remarks about Sudoku Squares". In: *The Mathematical Gazette* 90.519 (2006), pp. 425–430. ISSN: 00255572. URL: `http://www.jstor.org/stable/40378190` (visited on 03/11/2025).

[10] *Last remaining cell - Sudoku technique*. Accessed: 2025-02-18. URL: `https://sudoku.com/sudoku-rules/last-remaining-cell/`.

[11] Merriam-Webster Dictionary. *Sudoku*. Accessed: 2024-11-15. n.d. URL: `https://www.merriam-webster.com/dictionary/sudoku`.

[12] Peter Norvig. *Solving Every Sudoku Puzzle*. Accessed: 2025-02-18. 2006. URL: `https://www.norvig.com/sudoku.html`.

[13] Jason Rosenhouse and Laura Taalman. *Taking sudoku seriously: The math behind the world's most popular pencil puzzle*. Oxford University Press, 2012.

[14] Ed Russell and Frazer Jarvis. "Mathematics of sudoku II". In: *Mathematical Spectrum* 39.2 (2006), pp. 54–58.

[15] David Smith. "So you thought Sudoku came from the Land of the Rising Sun ..." In: *The Guardian* (May 2005). Accessed: 2025-01-03. URL: `https://www.theguardian.com/media/2005/may/15/pressandpublishing.usnews`.

[16] *What is a Shidoku? The 4×4 Sudoku Variant*. Accessed: 2025-01-19. URL: https://masteringsudoku.com/shidoku/#:~:text=A%20shidoku%20is%20a%20sudoku,meaning%20'four'%20in%20Japanese..