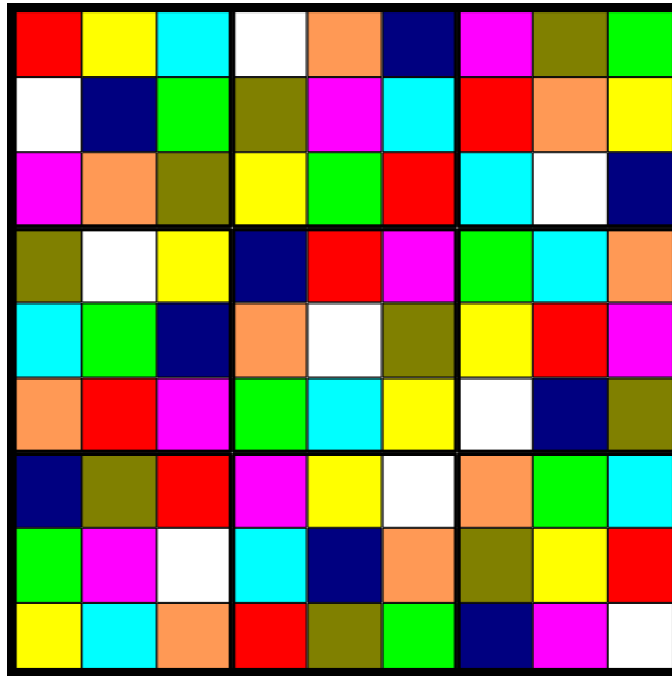# Playing Fair at Sudoku

Joshua Cooper

USC Department of Mathematics

Rules: Place the numbers 1 through 9 in the 81 boxes, but do not let any number appear twice in any row, column, or 3×3 "box".
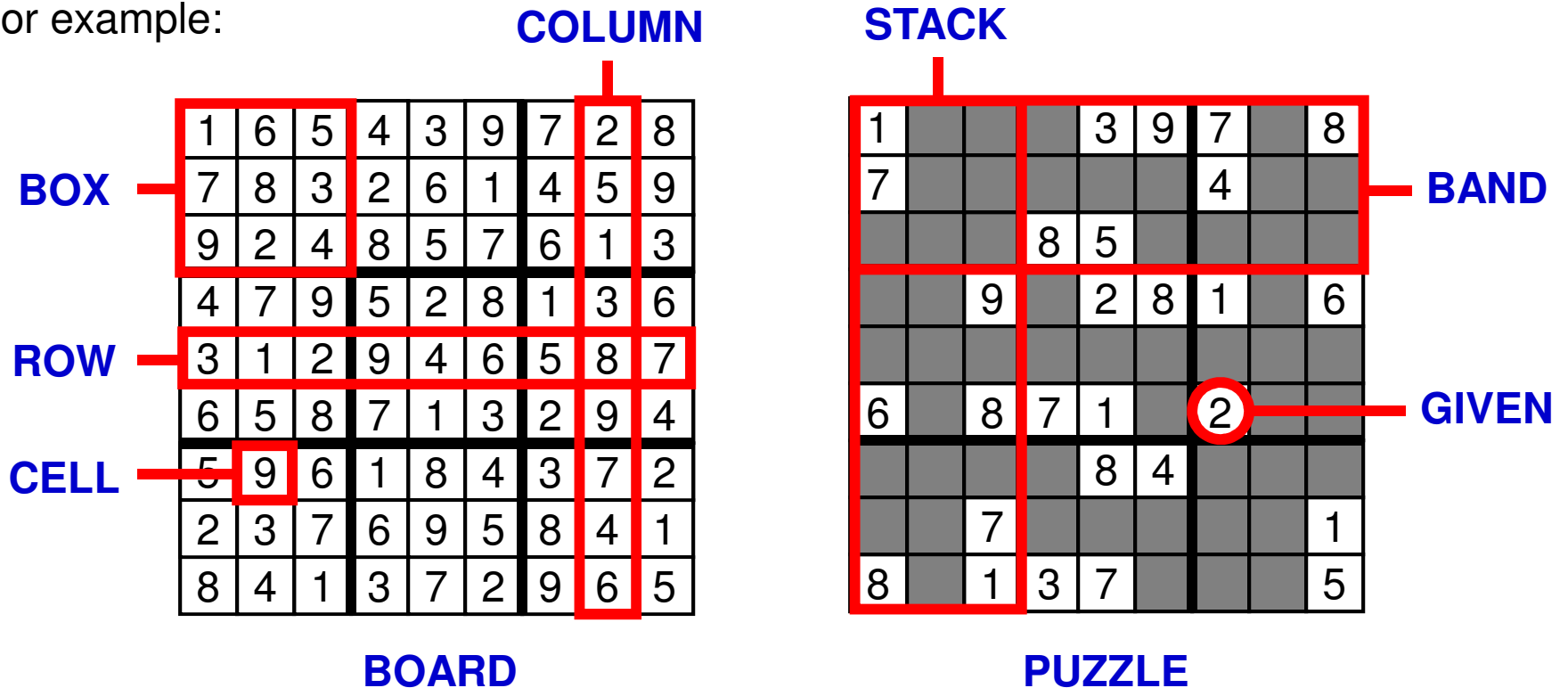
You start with a subset of the cells labeled, and try to finish it.

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

A Sudoku puzzle designer has two main tasks:

1. Come up with a board to use as the solution state.

2. Designate some subset of the board's squares as the initially exposed numbers ("givens").

For example:

We're going to focus on task #1: *How to choose a "fair" Sudoku board?*

For a Sudoku puzzle, i.e., a set of givens, to be "fair", it must have two properties:

1. It has a solution. (*Solvability*)

2. There is only one solution. (*Uniqueness*)

Question: What is the fewest number of givens in a fair puzzle?

Possible solution ("Brute Force"):

1. Enumerate all possible sets of givens.

2. Check each one to see if it is solvable.

3. Check the solvable ones to see if they are unique.

4. Count up the number of givens in the smallest uniquely solvable puzzle, and output the minimum such number.

Why Brute Force Is Impractical:

       1. Enumerate all possible sets of givens.

With 81 cells, there are $2^{81} \approx 2.4 \cdot 10^{24}$ sets of cells one could fill in.

Actually, the situation is even worse, because we have 9 options for the contents of each cell. That means a total number

$$1 + 9 \cdot 81 + 9^2 \cdot \binom{81}{2} + 9^3 \cdot \binom{81}{3} + \ldots + 9^{80} \cdot \binom{81}{80} + 9^{81} \cdot \binom{81}{81}$$

of possible sets of givens.

"81 choose 3" = the number of ways to choose 3 objects from a collection of 81

$$= \frac{81!}{78! \cdot 3!} = \frac{81 \cdot 80 \cdot 79}{3 \cdot 2 \cdot 1} = 1080.$$

Why Brute Force Is Impractical:

    1. Enumerate all possible sets of givens.

With 81 cells, there are $2^{81} \approx 2.4 \cdot 10^{24}$ sets of cells one could fill in.

Actually, the situation is even worse, because we have 9 options for the contents of each cell. That means a total number

$$1 + 9 \cdot 81 + 9^2 \cdot \binom{81}{2} + 9^3 \cdot \binom{81}{3} + \ldots + 9^{80} \cdot \binom{81}{80} + 9^{81} \cdot \binom{81}{81}$$

of possible sets of givens.

"**N** choose **K**" = the number of ways to choose **K** objects from a collection of **N**

$$= \frac{N!}{(N-K)! \cdot K!} = \text{BIG.}$$

Why Brute Force Is Impractical:

1. Enumerate all possible sets of givens…

With 81 cells, there are $2^{81} \approx 2.4 \cdot 10^{24}$ sets of cells one could fill in.

Actually, the situation is much worse, because we have 9 options for the contents of each cell. That means a total number

$$1 + 9 \cdot 81 + 9^2 \cdot \binom{81}{2} + 9^3 \cdot \binom{81}{3} + \ldots + 9^{80} \cdot \binom{81}{80} + 9^{81} \cdot \binom{81}{81}$$

of possible sets of givens.

By the Binomial Theorem,

$$\sum_{j=0}^{81} 9^j \binom{81}{j} = (1+9)^{81} = 10^{81},$$

which is approximately the number of atoms in the observable universe.

Let's be a little smarter about this…

1. Enumerate all sets of 81 givens, and if a uniquely satisfiable puzzle is found, enumerate all sets of 80 givens, and if a uniquely satisfiable puzzle is found, enumerate all sets of 79 givens…

In fact, we can start much lower than 81, since there are many uniquely satisfiable puzzles known with fewer than 81 givens.

Indeed, there are uniquely satisfiable puzzles known which have only 17 givens.

Gordon Royle has compiled a list of 49151 (!) *inequivalent* ones at:

http://mapleta.maths.uwa.edu.au/~gordon/sudokumin.php

What does it mean for two Sudoku boards/puzzles to be *equivalent*?

Two boards are considered *equivalent* if it is possible to transform one into the other by a sequence of operations of the form:

1. Permuting the rows and columns of each band/stack (X $3!^6$)

2. Permuting bands I, II, and III, and and stacks A, B, and C (X $3!^2$)

3. Permuting the numbers/colors (X $9!$)

This generates a *group* of 3,359,232 different possible operations.
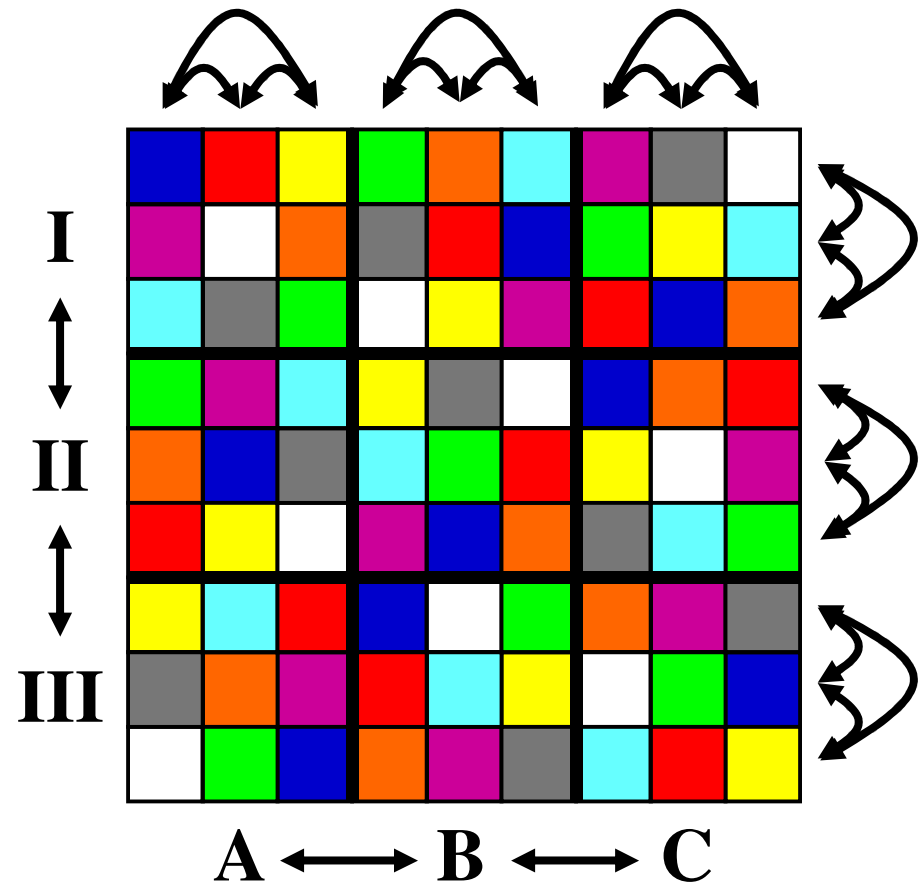
So, start with 16 givens:

       1. Enumerate all sets of 16 givens…

How many such sets are there?

$$9^{16} \cdot \binom{81}{16} \approx 6.22 \cdot 10^{31}$$

It would be silly to look at all of these, though:

       1. We can rule out anything that has two of the same symbol in any column, row, or box.
       2. Once we examine one, we don't have to look at all the ones equivalent to it.

By what factor does each of these reduce the number of sets to examine?

Effect of #1: This is nearly impossible to compute exactly, so we *Poissonize*: pretend as if each cell "turns on" with probability 16/81, and then chooses a number 1 through 9 uniformly at random.

Probability that two cells in the same row get turned on and set to the same number =

= Pr(exactly 2 cells get turned on and they are set to the same number) +
Pr(exactly 3 cells get turned on and at least 2 are set to the same number) +

$$\bullet$$
$$\bullet$$
$$\bullet$$

Pr(exactly 9 cells get turned on and at least 2 are set to the same number)

$$= \binom{9}{2} \cdot \left(\frac{16}{81}\right)^2 \cdot \left(\frac{65}{81}\right)^7 \cdot \frac{1}{9} + \binom{9}{3} \cdot \left(\frac{16}{81}\right)^3 \cdot \left(\frac{65}{81}\right)^6 \cdot \left(1 - \frac{9 \cdot 8 \cdot 7}{9^3}\right)$$

$$+ \binom{9}{4} \cdot \left(\frac{16}{81}\right)^4 \cdot \left(\frac{65}{81}\right)^5 \cdot \left(1 - \frac{9 \cdot 8 \cdot 7 \cdot 6}{9^4}\right) + \binom{9}{5} \cdot \left(\frac{16}{81}\right)^5 \cdot \left(\frac{65}{81}\right)^4 \cdot \left(1 - \frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{9^5}\right)$$

$$+ \cdots + \binom{9}{9} \cdot \left(\frac{16}{81}\right)^9 \cdot \left(\frac{65}{81}\right)^0 \cdot \left(1 - \frac{9!}{9^4}\right)$$

= 0.135469925561528... $\longleftarrow$ **Thanks, SAGE!**

Probability that **no** two cells in the same row get turned on and set to the same number:

$$= 1 - 0.135469925561528\ldots = 0.864530074438472\ldots$$

Probability that no two cells in **any** row get turned on and set to the same number:

$$= (0.864530074438472\ldots)^9 = 0.269786958\ldots$$

Probability that no two cells in any **row, column, or box** get turned on and set to the same number:

$$\approx 0.269786958\ldots^3 = 0.0196364445\ldots$$

Approximate total number of inequivalent configurations of 16 non-conflicting givens:

$$6.22 \cdot 10^{31} \cdot 0.0196364445 / 3359232 \approx 3.64 \times 10^{23}$$

Still **way** too big.

Even if we could enumerate all of these, and *even if* we knew how to generate a list of one representative of each equivalence class (= orbit under the Sudoku group)…

2. Check each one to see if it is solvable.
3. Check the solvable ones to see if they are unique. } Use *backtracking*.

What about lower bounds?

State of the art: At least 8 givens are needed.

Proof: Suppose fewer givens are provided. Then at least two numbers do not appear among the givens. Those two numbers could be switched in any completion of the board, so the solution is not unique.

□

Theorem (C' tomorrow?): At least **9** givens are needed.

Proof: Suppose only 8 givens are provided. First of all, note that no two of them are the same number, since otherwise there would be at least two numbers not appearing among the givens. Then, in any completion of the board, those two numbers could be swapped.

Next, set up a bipartite graph with the stack numbers and the band numbers as the color classes:

STACK 1 ○————○ BAND 1

STACK 2 ○————○ BAND 2

STACK 3 ○————○ BAND 3
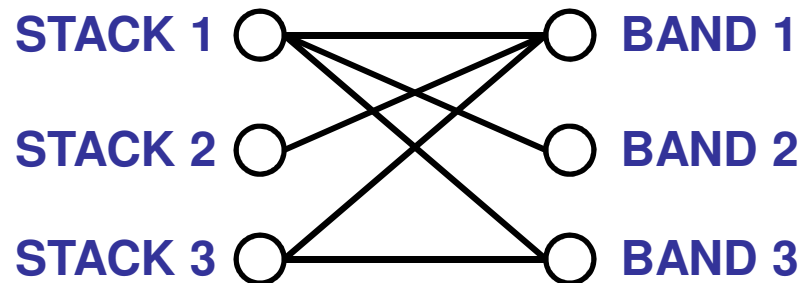
Now, put an edge from **STACK A** to **BAND B** if the box shared by them has a given in it.

When is there a *perfect matching*?

Definition. A perfect matching in a bipartite graph is a set of edges that connect each element of one color class to exactly one element of the other.

Theorem (Hall 1935): There is a perfect matching if and only if every subset of vertices in one color class has at least as many neighbors in the other color class as its size.

So, if there *isn't* a perfect matching, then there has to be either (a) one band containing no givens, (b) two bands with givens in only one stack, or (c) three bands with givens in only two stacks.

(a)	(b)	(c)
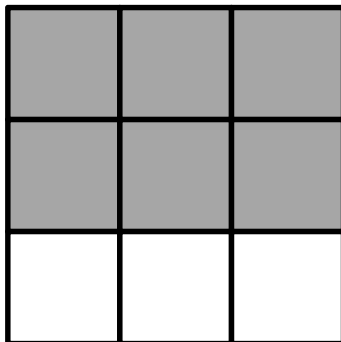
Permute the rows in the bottom band.	Not so easy.	Permute the columns in the left-hand band.
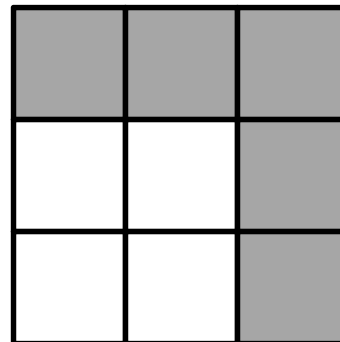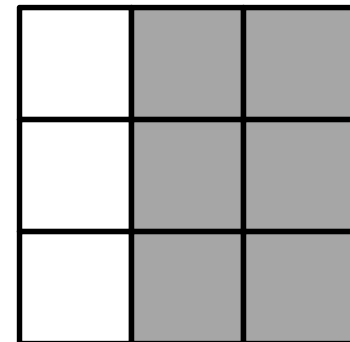
Case 1
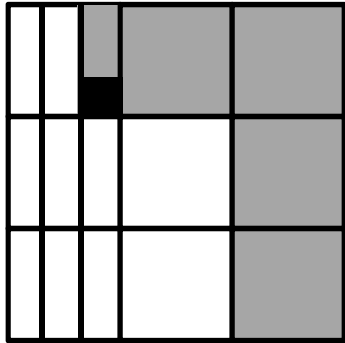
Swap these two columns.

Case 2
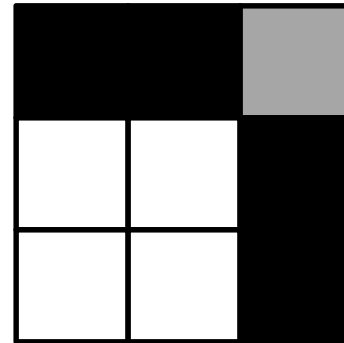
Since all four black boxes contain at least two givens, and there are only 8 givens in all, the gray square is empty as well.
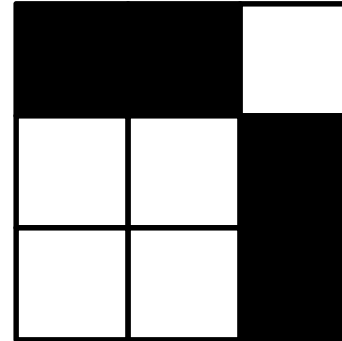
Case 2
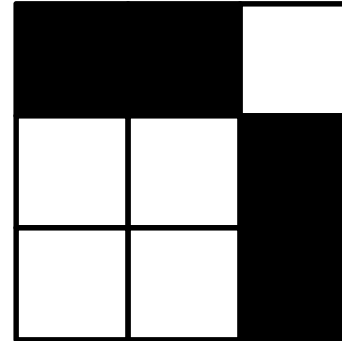
Since all four black boxes contain at least two givens, and there are only 8 givens in all, the gray square is empty as well.
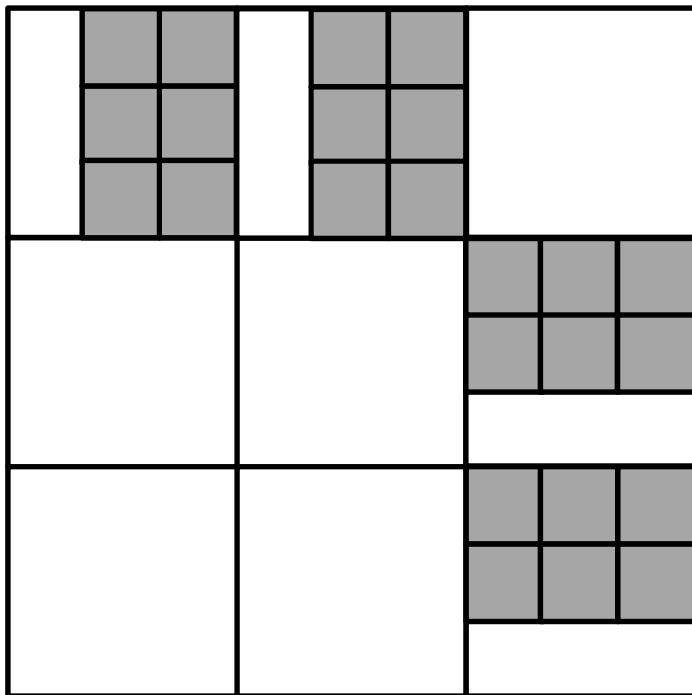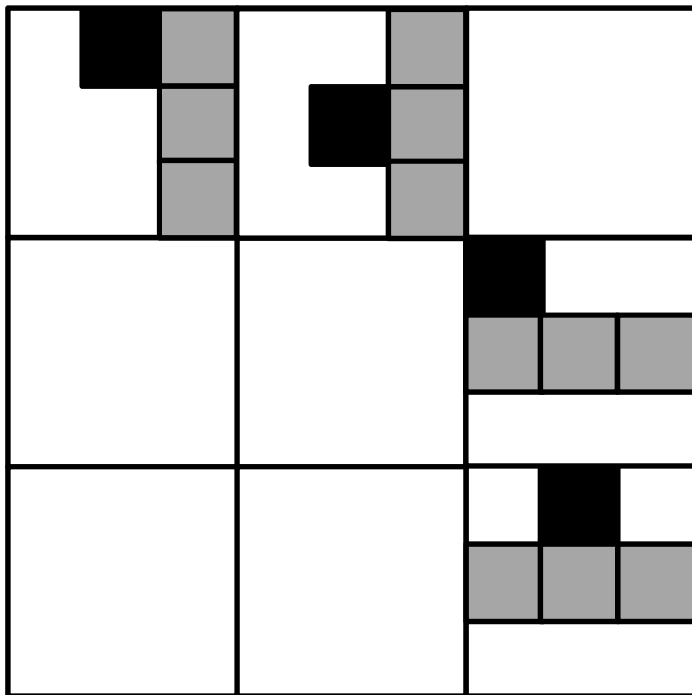
In fact, each black square has to have *exactly* two givens in it. So the picture is now this.

Case 2

Since all four black boxes contain at least two givens, and there are only 8 givens in all, the gray square is empty as well.

By permuting columns and rows, we can even assume more.

There are only $3^4 = 81$ choices for where The remaining four givens go. $\square$

If there *is* a perfect matching, then we may permute rows, columns, stacks, and bands so that the result looks like:
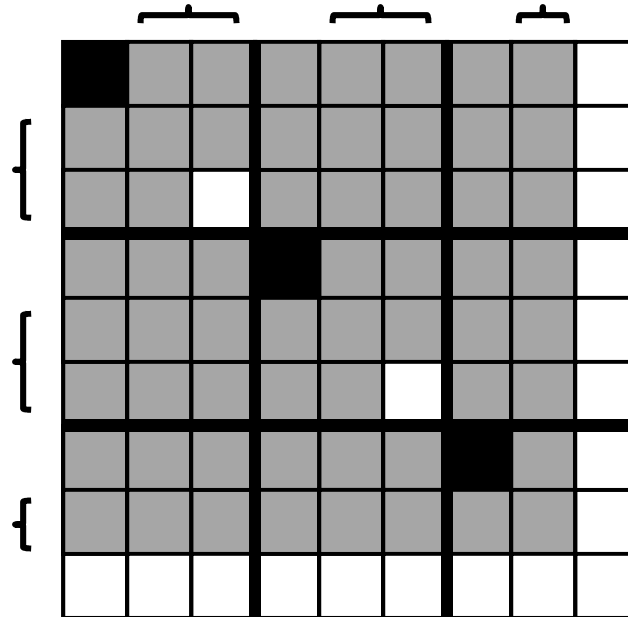
Out of the remaining 59 cells, 5 givens must be placed.

Each bracket must contain at least one given.

At least one of the boxes is empty, and there are essentially two ways to choose which one it is.

If there *is* a perfect matching, then we may permute rows, columns, stacks, and bands so that the result looks like:
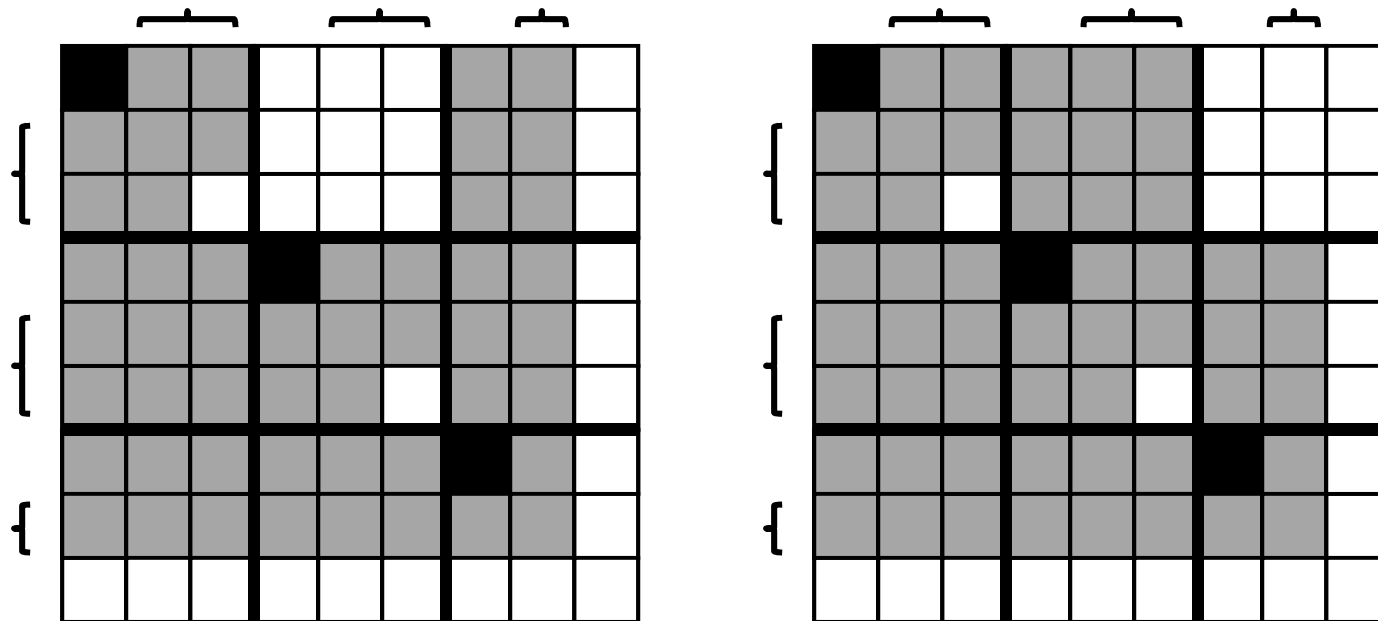
Out of the remaining 59 cells, 5 givens must be placed.

Each bracket must contain at least one given.

At least one of the boxes is empty, and there are essentially two ways to choose which one it is.

$$\binom{50}{5}+\binom{53}{5}+81=4{,}988{,}526$$

How to test if a puzzle is fair?

Define a graph Sud on the set of cells with a complete subgraph in each row, column, and box.
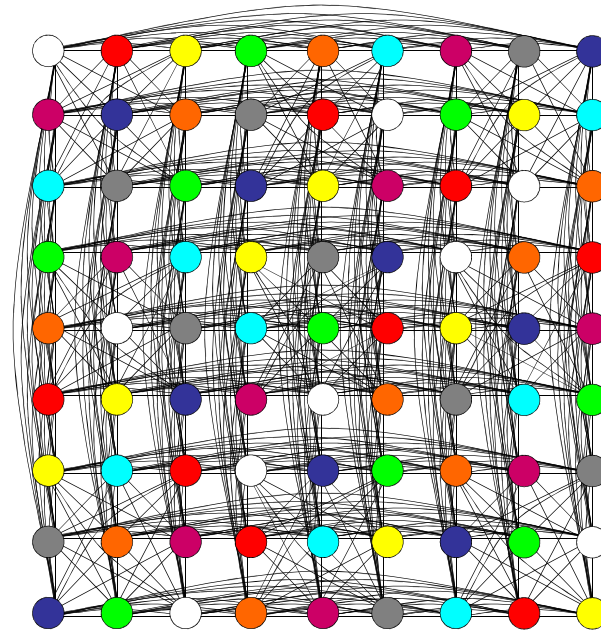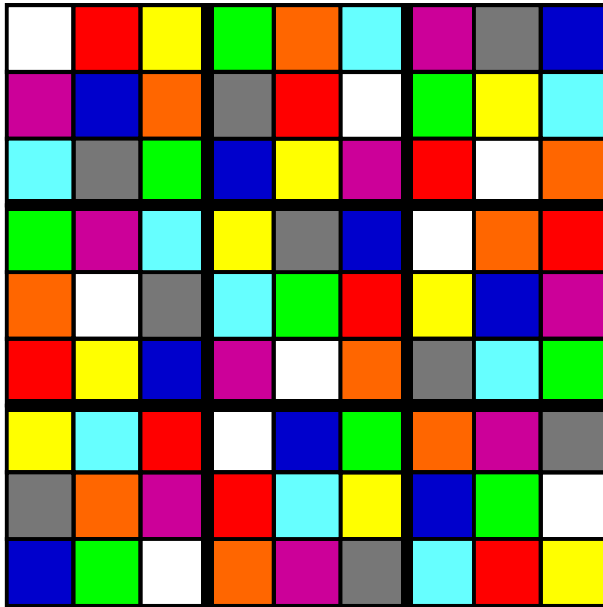

Definition. A graph G is said to be *k-colorable* if it is possible to assign *k* colors to the vertices in such a way that no edge has both its vertices colored the same.

Definition. The *chromatic number* $\chi(G)$ of a graph G is the smallest integer *k* so that G is *k-colorable*.

Definition. A graph G is said to be *uniquely colorable* if, other than permuting the colors, there is only one coloring of G with exactly $\chi(G)$ colors.
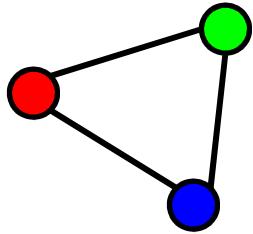
Uniquely Colorable
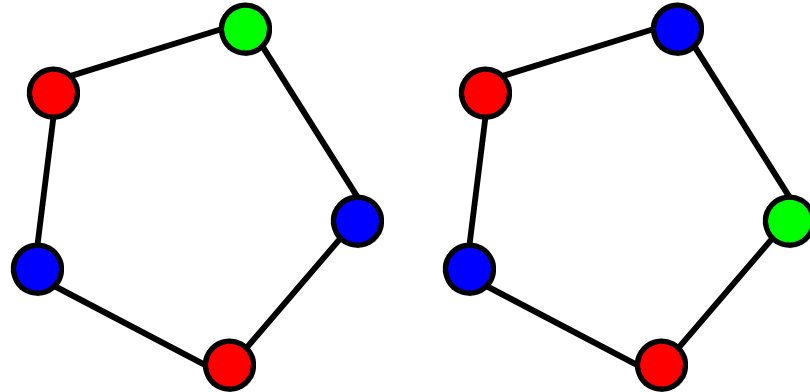
Not Uniquely Colorable

Theorem. At least 9 givens are required if, for all sets S of 8 givens, the graph obtained from Sud by adding a complete graph on the vertex set corresponding to S is *not* uniquely colorable.

Proof. A coloring of Sud is exactly the same thing as a solution to the puzzle. When a complete graph is added, we are simply requiring that the 8 givens be assigned different colors. We may assign numbers uniquely to the cells exactly when the graph is uniquely colorable.

The gap between 9 and 17 is still huge!

Open Problems:

1. Close the gap!

2. How many fair puzzles are there with *k* givens, 9 ≤ *k* ≤ 81?

3. What about the "generalized Sudoku board"? For example, 16X16:

Thanks!

P.S. If you are interested in doing some research, contact me at cooper@math.sc.edu.