

How To Solve Sudoku Puzzles

2nd Edition

Copyright © 2011, 2013 by Jim Scott, 1711 Habwood Lane, Henrico VA 23238 USA

Introduction

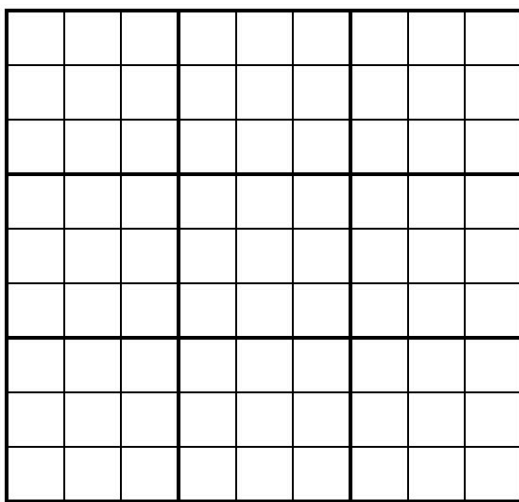
If you buy a sudoku book in a bookstore or on a newsstand, it's likely to be just a collection of puzzles, perhaps with a few hints about how to solve sudoku puzzles. You'll be lucky to find a practical discussion of solution methods, although some do exist. Check out the sudoku web sites, and you'll find mathematical treatises, discussions of arcane techniques, and descriptions of non-standard versions of sudoku.

My purpose here is to present a method that should be sufficient to solve any *standard* sudoku puzzle you're likely to encounter. It's not an *all-inclusive* list of sudoku-solving tools, and it's not a discussion of how to solve sudoku puzzles *faster* than other people. It's just a method for getting to a solution.

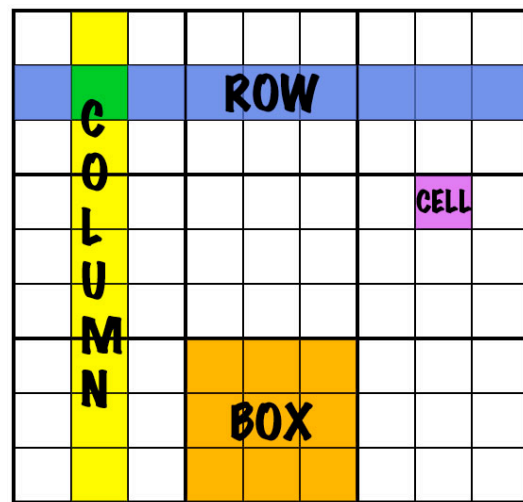
This discussion deals only with standard sudoku, the version that has a 9x9 grid and uses the digits 1-9. (If you're not sure what it means to solve a sudoku puzzle, read on; it will be made clear at the end of the next section.) My goal is to show you exactly what you need to know to solve even relatively difficult sudoku puzzles, while omitting the stuff that is of interest only to theorists.

Terminology - Part 1

There's no standard glossary of sudoku terms, but most of the following definitions are commonly used. This list includes only the terms that are needed for this discussion.



Grid



Cell/Row/Column/Box

Grid - The game diagram, divided into nine rows, nine columns, and 81 cells.

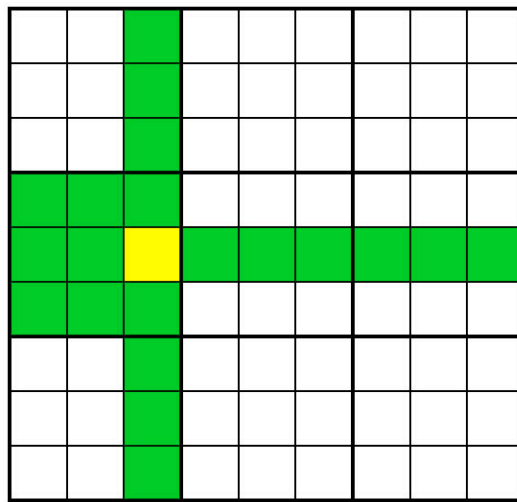
Cell - One of 81 squares that will (eventually) hold a single digit.

Row - A horizontal line of nine cells. There are nine rows.

Column - A vertical line of nine cells. There are nine columns.

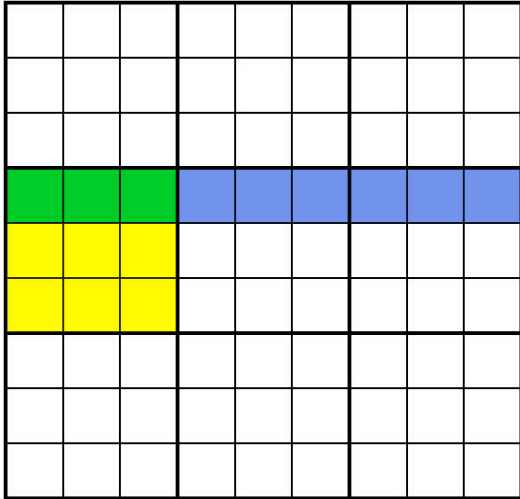
Box - A 3x3 group of nine cells, separated from other boxes by heavy lines. There are nine boxes.

House - A row, column, or box of nine cells. There are 27 houses.

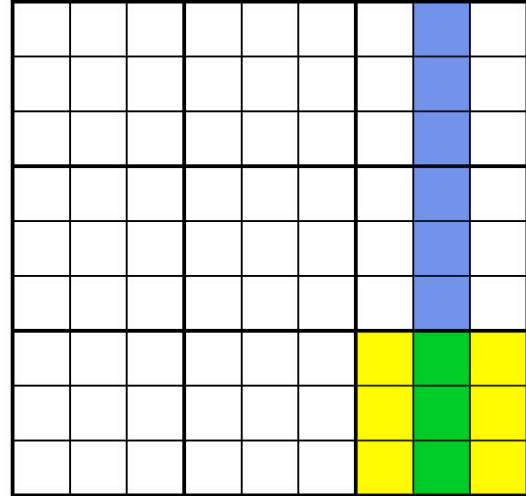


Neighbor - The 20 green cells are neighbors of the yellow cell.

Neighbor - A cell that shares a house with another cell. Any specific cell lies inside three houses, i.e., one row, one column, and one box. Within those three houses, the cell has 20 neighbors, namely, the other eight cells in its row, the other eight cells in its column, and four more cells in its box that aren't in its row or column.



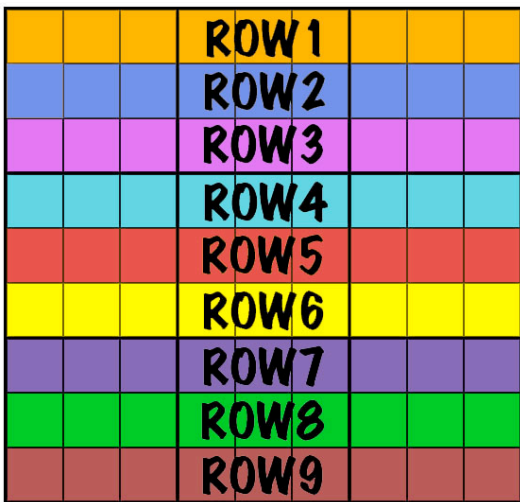
Boxrow - The blue row and the yellow box intersect at the green boxrow.



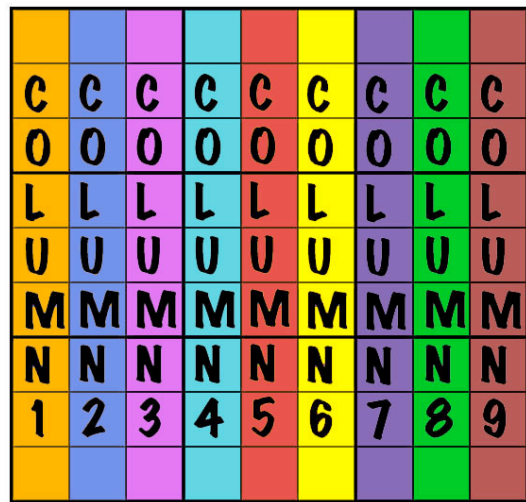
Boxcol - The blue column and the yellow box intersect at the green boxcol.

Boxrow - The intersection of a box and a row.

Boxcol - The intersection of a box and a column.



Row#



Col#

Row# - A notation used to specify a row. Row1 is the row at the top of the grid, and Row9 is at the bottom.

Col# - A notation used to specify a column. Col1 is the column at the left of the grid, and Col9 is at the right.

Terminology - Part 2

Naked Pair - A *naked pair* is two cells in the same house that have exactly the same two candidate digits, and *only* those two digits. When a naked pair exists, no other cell in the house, other than those two, can contain either of the two digits. In other words, you can examine each of the other seven cells in the house and eliminate any occurrence as a candidate of either of the two naked pair digits.

A naked pair will sometimes share two houses at once, for example a row and a box. In that case, you can eliminate the two digits as candidates anywhere they appear in the other (13) cells of the two houses.

37	35 79	8
4	2	6
1	579	37

Naked Pair - Before

37	59	8
4	2	6
1	59	37

Naked Pair - After

In the diagram above left, only Box1 is shown. Five cells have been established, and the other four have been penciled in. Since R1C1 and R3C3 (the yellow cells) both contain 37 (and no other digits), they are a naked pair within that box. Any other 3 or 7 penciled into cells in that box can be eliminated. In the diagram on the right the 3 and 7 have been eliminated from R1C2, leaving 59, and the 7 has been eliminated from R3C2, again leaving 59. As it happens, R1C2 and R3C2, the two cells containing 59 in the right diagram, are now a new naked pair. (This isn't always the case.) This naked pair exists within one boxcol, so they affect Col2 as well as Box1.

Naked Triple - A *naked triple* is three cells in the same house that share the same three candidate digits, or any two of the three. When a naked triple exists, no other cell in the house can contain any of those three digits; any occurrence of any of those three digits as candidates elsewhere in the house can be removed. As with a naked pair, it's possible for a naked triple to share two houses.

167	9	134 67
678	678	234 678
5	68	124 68

Naked Triple - Before

1	9	34
678	678	234
5	68	24

Naked Triple - After

In the box above left, 9 and 5 have been given or filled in for two of the cells, and the other seven cells have multiple candidates penciled in. The three yellow cells all contain only 6, 7, and/or 8, so they are a naked triple. Therefore, 6, 7, and 8 can be eliminated as candidates in the other cells. In the upper left cell, that leaves only one candidate, 1, so the 1 is filled in there. Now 1 can also be eliminated as a candidate from the remaining cells, leaving the box above right. Note that the right three cells now have become a new naked triple containing 2, 3, and 4. (This isn't always be the case.) This is a naked triple not only within the box, but also within the column that passes through those three cells (i.e., within a boxcol). The other six cells of that column are not shown here, but 2, 3, and 4 would be eliminated as candidates in those six cells.

Now see if you can find the naked triple in the following row:

25	3	259	26 78	1	26 78	4	25 89	259
----	----------	-----	----------	----------	----------	----------	----------	-----

The first, third, and ninth cells contain only the digits 2, 5, and 9 (or a subset, as in the first cell).

25	3	259	26 78	1	26 78	4	25 89	259
----	----------	-----	----------	----------	----------	----------	----------	-----

So those three cells, shown here in yellow, make up a naked triple.

25	3	259	678	1	678	4	8	259
----	----------	-----	-----	----------	-----	----------	----------	-----

Consequently, those three digits, 2, 5, and 9, are eliminated as candidates from other cells in the row. The 2678 in cells four and six becomes 678, and the 2589 in cell eight gets filled in as 8.

25	3	259	67	1	67	4	8	259
----	----------	-----	----	----------	----	----------	----------	-----

Since 8 is now established in the row, the 678 in cells four and six becomes 67. Note that this pair of 67s is now a naked pair; what's more, it lies within a single boxrow. Therefore, 6 and 7 can be removed as candidates from the other cells (not shown here) of the box that contains that boxrow.

Naked Quad - A *naked quad* is four cells in the same house that share the same four candidate digits, or any two or three of the four. Those four digits can be removed as candidates from the other five cells of the house. A naked quad cannot be shared by two houses. Useful naked quads are not common; an example of a non-useful naked quad is one that exists in a house whose other five cells are already established (given or filled in).

Crosshatching - *Crosshatching* is a process of deciding which cells in a particular house are eligible to contain a particular digit, by looking at locations where that digit is established in shared houses.

								3
					3	1		
					5			
9			3		4		1	
				9		8	2	
				5				

Crosshatching - Start

								3
						3	1	
						5		
9			3			4		1
					9		8	2
					5			

Crosshatching - Finish

For example, in the grid at left above, suppose we need to know which cell or cells in Box9 (green) might contain a 3. To do this, we look to see where any 3's are established in any of the rows and columns that intersect with Box9. (The diagram doesn't show any penciled-in digits because crosshatching uses only established digits. Also, Box1-Box2 and Box4-Box5 are not shown because they don't matter when crosshatching Box9.) In the right grid above, Col7, Col9, and Row7 have all been colored yellow because they each contain a 3. This is the process of crosshatching. In Box9, only cells R8C8 and R9C8 are left untouched. R8C8 is already occupied (it contains an 8), so R9C8 (orange) is the only possible location for a 3 in Box9.

Crosshatching is not usually done exhaustively, i.e. for each of the nine digits in each of the nine boxes, which would be 81 separate crosshatching procedures (minus any given digits). What's more, it typically does *not* pin down a single cell where the digit belongs. However, crosshatching can be used selectively, when you really do want to know where a particular digit might occur in a particular box, as described in the next section.

Solution Process

With all of that background out of the way, we're now ready to describe a procedure that should lead to the solution of any standard sudoku puzzle that the typical puzzler is likely to encounter. In practice, you would probably not follow this procedure slavishly. I.e., you wouldn't necessarily do each step in the order described here; there's nothing wrong with skipping ahead in the process. If you get stuck, go back and proceed systematically.

Step 1 - Initial Scan with Selective Crosshatching

Start by examining the top three boxes, Box1-Box3. For each of the nine digits, see if there's any digit that occurs in exactly two of those three boxes. Use crosshatching to find where that digit might belong in the third box. If you can narrow it down to one cell, fill in the digit. Then continue the initial scan.

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

For example, in the grid above, look at Box1-Box3 (green). The digits 1, 2, 5, 6, and 9 occur once each, so we're not interested in them at this point. The digits 4 and 7 don't occur at all in Box1-Box3. But the digits 3 and 8 occur twice each, so we'll look at each of them.

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

Continuing the example, the 3 occurs in Box1 and Box3, and we want to know if we can figure out where to put a 3 in Box2. Using crosshatching, we eliminate Row1 (because of the 3 at R1C1), Row3 (thanks to the 3 at R3C9), and Col4 (because of the 3 at R8C4). Given that R2C6 is occupied (contains a 6), the 3 in Box2 must go into cell R2C5 (orange).

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

Having filled in that 3, we next use crosshatching to see where an 8 might go in Box1. Again crosshatching eliminates all but two cells, one of which is occupied. The 8 must go into R1C3.

Once Box1-Box3 have been scanned for all nine digits, continue for Box4-Box6, then Box7-Box9.

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

Continuing the example, we move on to Box4-Box6. We find one 1, no 2, one 3, no 4, two 5s, one 6, one 7, one 8, and three 9s. So we're only interested in the 5s, which occur in box 4 and box 6. We use crosshatching to see where a 5 might go in box 5, and we find three possibilities, R4C4, R4C5, and R4C6 (orange). We cannot place the 5 in box 5 at this point. Scanning Box7-Box9 is also unproductive.

After finishing with Box1-Box3, Box4-Box6, and Box7-Box9, go *vertical* with Box1/Box4/Box7; then Box2/Box5/Box8; and finally Box3/Box6/Box9, i.e., Col7-Col9.

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

In the example, processing Box1/Box4/Box7 produces no progress, and ditto with Box2/Box5/Box8. When we get to Box3/Box6/Box9, we find two 3s and two 9s. Crosshatching the 3s reveals that a 3 can be filled in at R9C8 (orange).

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

Finishing the example, we crosshatch the 9s in Box3 and box 6 to find that the 9 in Box9 must go into cell R8C8.

The initial scan will typically fill in at least a few cells, depending on how difficult the puzzle is. If you feel it's worth your time, you can repeat the initial scan until it produces no new information. The reason is that any newly-established digit could affect the result of further scanning. In the last example above, if we had crosshatched the 9s first, followed by the 3s, we wouldn't have known where to put the 9 in Box9. Can you see why? It's just a matter of luck. But if you repeat Step 1 (Initial Scan with Selective Crosshatching) for all 27 houses until an entire scan produces no further results, then you can be sure you won't have missed anything.

Step 2 - Pencil In All Candidates

Once you've completed Step 1, thereby reaping the low-hanging fruit, you're ready for the nitty-gritty work of penciling in each cell. At the end of this step, no cell will be empty.

For each empty cell, determine all of its candidates and pencil them in. You might want to start with the empty cells whose neighbors have lots of established digits, so that there may be relatively few candidates. But the order doesn't really matter; even a random order is okay.

For each cell, look at all the digits that occur in its 20 neighbor cells. **Any digit that is not established in any neighbor is a candidate for that cellⁱⁱ.** Using whatever notation you like, write all of the candidates into that cell. (Solving a sudoku puzzle can be seen as a process of reducing the number of candidates in each cell to one.)

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

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

For example, in the diagram at left above (the same grid used in Step 1), suppose you're ready to pencil in cell R1C2 (orange). Look at all of the cells that neighbor R1C2, namely Row1, Col2, and Box1 (yellow in second diagram).

3	⁴⁶ ₇₉			2	1			
					6	8		
	5				8	9		3
9						3		
7	8		6		9		1	5
		5						9
5		1	8				6	
		4	3					
			1	6				4

3	⁴⁶ ₇₉			2	1			
					6	8		
	5				8	9		3
9						3		
7	8		6		9		1	5
		5						9
5		1	8				6	
		4	3					
			1	6				4

In those neighbor cells, the digits 1, 2, 3, 5, and 8 are established, so the candidates in cell R1C2 are the other digits, namely 4, 6, 7, and 9. In the diagram on the left above, cell R1C2 has been penciled in. Now we move on to cell R1C3 (orange) and its neighbors (yellow). The established digits are 1, 2, 3, 4, and 5. *The digits penciled in at R1C2 don't matter in this process.*

3	46 79	67 89		2	1			
					6	8		
	5				8	9		3
9						3		
7	8		6		9		1	5
		5						9
5		1	8				6	
		4	3					
			1	6				4

So the candidates for R1C3 are 6, 7, 8, 9.

If it turns out that there's only one candidate for a cell, congratulations! Go ahead and fill it in.

In penciling in the candidates, there are different types of error that you can make.

Type 1 Error (Duplicate Established Digit) - A digit is filled into a cell in a house that already has that same digit established in another cell. This error is almost certain to be fatal unless you catch it quickly.

Type 2 Error (Omitted Candidate) - A legitimate candidate is omitted from a cell. In this case, you may (eventually) find that you cannot solve the puzzle.

Type 3 Error (Superfluous Candidate) - A digit is included as a candidate in a cell, but it isn't really a candidate in that cell. This type of error shouldn't hurt you permanently; sooner or later you'll find a reason to eliminate that digit as a candidate for that cell. However, this error can really slow your progress. (See "Step 3 - Error Checking", below, for a way to deal with this issue.)

Once Steps 1 and 2 are finished, every cell is either a given, or has been filled in, or has all of its candidates penciled in.

Step 3 - Error Checking

This step is not strictly necessary, but it's highly recommended.

Scan each row, then each column, then each box. (The order doesn't matter, as long as you're systematic.) For each of the 27 houses, look first at which digits have already been established (given or filled in). Then examine each of the other cells of this house, the ones that are penciled in, looking for any instances where one of the established digits is also a

candidate in a penciled-in cell. In other words, look for instances of Type 3 Errors (Superfluous Candidate).

8	1	59	2	37	4	367	678	59
----------	----------	----	----------	----	----------	-----	-----	----

For example, in the row above, the digits 1, 2, 4, and 8 have been established. Now scan the other cells, the ones that are penciled in, looking for any occurrence of any of those four digits. All is well except in the next-to-last cell, where we find 678 penciled in. One of those digits, 8, has already been established (in the leftmost cell).

8	1	59	2	37	4	367	67	59
----------	----------	----	----------	----	----------	-----	----	----

So digit 8 should be erased from the next-to-last cell.

Continue this error check for all 27 houses. You may be surprised at how many errors of this type you find. The important part is that by finding them now, before beginning Step 4, you'll save yourself a lot of time.

Step 4 - Final Scan

Now you're ready for the Final Scan, the tofuⁱⁱⁱ and potatoes of the solution process. Again, scan each row, each column, and each box, using any systematic scan you prefer that covers all 27 houses. For each house, do each of the following checks.

If at any point you find that all candidates but one have been eliminated from a cell, fill in that cell with the one remaining candidate. Then eliminate that digit as a candidate from all of the neighbors of that cell, i.e., from all other cells in its row, in its column, *and* in its box.

- 1) **Eight Established Digits** - Look for a house that has established digits in exactly eight cells. Fill in the ninth cell with the only possible digit.
- 2) **Naked Pair** - Look for a naked pair, two cells that have the same two candidates. If you find a naked pair, then those two candidates can be eliminated from the other seven cells in that house. If the two cells of the naked pair fall into one boxrow or boxcol, then *two* houses benefit, not just one.
- 3) **Naked Triple** - Look for a naked triple, three cells that share the same three candidates, or any two of the three. If you find a naked triple, then those three candidates can be

eliminated from the other six cells in that house. Again, if all three cells are in one boxrow or boxcol, two houses are affected.

- 4) **Naked Quad** - Look for a naked quad, four cells that share the same four candidates, or any two or three of the four. If you find a naked quad, then those four candidates can be eliminated from the other five cells in that house.
- 5) **Digit Count** - For each of the nine digits, examine all nine cells of the house and determine how many times that digit occurs, either given, filled in, or as a candidate. Initially, you only need to see whether the digit occurs once or more than once. If it does not occur at all, then a Type 2 Error (Omitted Candidate) has been made; at least one legitimate candidate has been missed in Step 2, and you may be in trouble.
 - a) **Sole Candidate** - If the digit occurs exactly once, and it's in a penciled-in cell, then it has just become established (all of the other candidates for that cell are eliminated); fill it into that cell. Also, eliminate the newly-established digit from the other two houses that the cell occupies.
 - b) **Superfluous Candidate** - If the digit occurs more than once, *both* as a candidate *and* as an established digit, then a Type 3 Error (Superfluous Candidate) has been made. (Step 3, Error Checking, tries to eliminate these errors. But they can still occur, either because Step 3 missed one, or because the error was created earlier in Step 4, after Step 3 was finished.) But that's okay; simply eliminate the digit in question where it appears as a candidate. (If the digit is *established* in more than one cell, then a Type 1 Error (Duplicate Established Digit) has occurred, and the game is over unless you can find the cause of the error and correct it.)
 - c) **Candidate Digit Confined to One Boxrow/Boxcol** - If the digit is not established anywhere in that house, but occurs more than once as a candidate, determine whether all of the occurrences fall into one boxrow or one boxcol. **If they do, eliminate that digit as a candidate from any of the other six cells of the intersected house.**

236 79	23 79	379	5	1	8	39	4	369
369	39	5	7	24	24	1	69	8
1	4	8	36	69	39	7	2	5

For example, the diagram above shows only Box1-Box3. Let's say you're currently scanning boxes. When you get to Box3 (yellow), you look for naked pairs and find none. Next you look for naked triples, and you find one, the three cells with 39, 369, and 69 penciled in. However, it makes no difference because all of the other cells in the box are established. The naked triple doesn't lie within one boxrow or one

boxcol, so it doesn't affect any intersecting row or column. Of course, there's no naked quad, and so on.

Continuing the example, finally you look for a digit that exists more than once as a (penciled-in) candidate within Box3, and for whom all occurrences fall within one boxrow or boxcol. This does not apply for 6 or 9, but both occurrences of 3 (as a candidate in Box3) are in the top boxrow. Therefore, 3 must eventually be established in the top boxrow of Box3.

236 79	23 79	379	5	1	8	39	4	369
369	39	5	7	24	24	1	69	8
1	4	8	36	69	39	7	2	5

But 3 must also lie somewhere in Row1 (now also yellow), and we know now that it must be one of those two candidate 3s (red) in the top boxrow of Box3. Therefore, the other occurrences of candidate 3s in Row1 (blue) should be eliminated.

26 79	279	79	5	1	8	39	4	369
369	39	5	7	24	24	1	69	8
1	4	8	36	69	39	7	2	5

The result is shown in the diagram above. We've eliminated 3 as a candidate from three cells, and that's definite progress.

Repeat the Final Scan as many times as necessary. Eventually, it becomes a chase of filling in a cell, eliminating the neighboring candidates that that rules out, finding that some of those neighbor cells can now be filled in, and so on.

Tips

When penciling in or filling in, write each digit neatly and consistently. This will allow you to perceive each digit quickly, thereby speeding up the scanning process.

When you're looking for naked pairs, look only at cells that have exactly two digits penciled in. When you're looking for naked triples, the only cells to examine are those with exactly two or three digits penciled in. (Note that a naked triple can consist of three cells that

contain exactly two digits each; e.g., 57, 59, 79.) And naked quads, of course, involve only cells with two, three, or four candidates each.

As already mentioned, if a house already has exactly five cells established, the remaining four cells must contain a naked quad. But it would be a useless naked quad, because there would be no candidates to be eliminated. Likewise, a house with exactly six cells established necessarily contains a naked triple, and a house with exactly seven cells established contains a naked pair. But in either of these two cases, if the naked pair or triple falls within one boxrow or boxcol, then it's shared by a second house, and there might be candidates to be eliminated in that house.

Whenever you fill in a cell (assuming you've finished Step 2, Penciling In), you might want to take the time to verify that each of the three houses that include that cell has all of the digits 1-9, either established or as candidates. If not all the digits 1-9 appear in a house, you must have a Type 1 Error (Duplicate Established Digit) or a Type 2 Error (Omitted Candidate). In this case, recheck the cell that you just penciled in. If you can find the problem there, then you might recover. You might have to backtrack more than one step to find the problem, but going more than two steps back is difficult. If you can't correct the error, then you're dead in the water. You must either start over or scrap the puzzle. How much erasing are you willing to do?

As you become familiar with all of these processes, you can merge Steps 2 and 4 (Penciling In and Final Scan). For example, in penciling in the cells of a row, you may recognize the existence of a naked triple in that row. You can then pencil in the rest of the empty cells in that row, omitting any further occurrences of the three digits of the naked triple.

In fact, if you feel like it, you can just randomly perform the various activities of Steps 2 and 4. But if you ever feel lost, first make sure that all cells are penciled in, and then begin a Final Scan (preceded, if you wish, by Step 3, Error Checking).

If you ever complete an entire Final Scan without making any progress at all, there's probably at least one Type 3 Error (Superfluous Candidate) somewhere. So redo Step 3, Error Checking. If you find a digit that occurs both as an established digit in a house and as a candidate in another cell of that house, then eliminate the candidate (but leave the established digit, of course). Then do a complete Final Scan process on each of three houses that include the affected cell.

Conclusion

That's all it takes to solve an ordinary sudoku puzzle. I know this method is not sufficient for *all* sudoku puzzles, but I've used it many times on the hardest sudoku puzzles I normally encounter (rated *Evil* on the web site www.websudoku.com), and I've never been permanently stymied by any of them. Either I've solved the puzzle, or I've found my attempted solution to be inconsistent, meaning that I made an error and wasn't willing to start over.

If the method were not sufficient, one would eventually find that there are still multiple cells with multiple candidates, and that applying the method fails to eliminate any further candidates. There was a time when I not infrequently found myself unable to make further progress on a sudoku puzzle. I put each of these aside, and came back to them months later. I was still using the same methods, but was now able to solve them, simply because I had become more observant with practice. In the end, the key is practice - plenty of it.

Footnotes

ⁱ Why "naked"? You might well ask. A comprehensive discussion of sudoku-solving tools

ⁱⁱ It's not strictly true that any digit that is not established in a neighbor of a cell must be a candidate in that cell, but the statement is good enough for the penciling-in process. As you gain experience and begin to combine steps, you might, for example, notice that the row containing the cell in question contains a naked pair. The two digits of that naked pair can be eliminated as candidates, even though they aren't established as such.

ⁱⁱⁱ Read "*meat* and potatoes" if you like; but I'm a vegetarian.