

Sudoku Rater Instructions

Thank you for your willingness to assist us in our effort to assess Sudoku Problem Solving Strategies.

We conducted an experiment in which participants were introduced to Sudoku, then received instructions in a specific strategy for filling in a digit in a Sudoku grid. Each participant was then given 25 practice problems with detailed feedback on the reasons why any incorrect responses were wrong and with the requirement to try again on each problem until success, followed by 64 additional problems where the feedback only indicated if their response was correct or not, and they only had one try to give a correct answer. The participants then solved one more puzzle, and while it was still visible, completed a questionnaire, which included the question below:

“Explain as clearly as possible the steps you went through to choose your answer. Please be as detailed as possible so that someone else could replicate your strategy by following your response.”

You have gone through the experiment yourself and have been able to do well in solving the presented problems, and you have answered the question above, as well as some additional questions, yourself. We now ask that you score the answers that participants gave to the above question, using a scoring rubric we have created. Initially, you will be scoring the responses of 20 participants. To be sure you understand the instructions and classification scheme, we will be checking your classifications of these 20 participants. We will then meet to discuss any clarifications and adjustments that need to be made before proceeding to classify the remaining participants' responses. Assuming our collective judgment is that the classification scheme is clear and we are in agreement about how it should be applied to the 20 training examples, you will then be asked to rate the entire set of 168 participants. Your compensation for participation will be \$30/hour, including the time you spent doing the experiment, reading the instructions, doing the first 20 ratings, and then doing the full set of ratings.

Detailed Response Classification Instructions

Please spend an hour making sure you understand these instructions and the Scoring Guide that is available at the **Guide** tab of the spreadsheet where you will record your ratings. Do this before proceeding to score the 20 responses we have provided. Once you start in on scoring the responses, you should feel free to consider each response against the class description, examples, and notes in the Scoring Guide as you consider each case. You can make tentative classifications that you finalize later throughout this process.

For each participant, you will see the puzzle the participant was given and their answer. The puzzles are provided in a PDF file while the participants' responses are in a Google Sheets file that you will use to complete your ratings. Please fill in only the white cells in the 'Ratings' tab, proceeding left to right. **Do not edit the cells in blue as that may break the program. Some of the columns are hidden; please do not unhide these columns as they contain information that should remain hidden until revealed by other columns.**

Your first task will be to determine if the participant's answer was correct or incorrect. You will determine this by looking at the grid the participant saw, and the answer they gave, relying on your own understanding of the logic of Sudoku and the constraints that apply to the puzzles. Once you enter your scoring of the participant's answer in the **Evaluation** column, the spreadsheet will show the correct scoring in the **Actual** column. This step is intended as an attention check. Please make sure you are attending to minimize your errors at this step. Correct any mistakes you make in your own scoring of the puzzle after verifying that you agree that your scoring was incorrect. If you think the Actual column is incorrect, contact ajhnam@stanford.edu immediately.

Next, you will see the participant's answer to the question probing their solution strategy, and you will be asked to answer three questions.

1. Focus on Prevalent Digits (PDs). In all of the puzzles, there were two 'prevalent digits', one of which was always the correct answer (also called 'target') and the other of which we call the 'distractor'. Your task here is to determine if the participant's response indicated that they focused on one or both of these digits by filling in the **PD** column in the spreadsheet. If they mentioned the target and distractor by name or otherwise made you think that they were aware of the fact that the answer was always one of the two digits that each occurred several times, you should choose the *Both* option. Alternatively the response may suggest a focus on just one of these digits. If you can determine that this was the target, you would choose the *Target* option in this case. If you can determine that this was the distractor, you would choose the *Distractor* option. You may rely on the participant's answer to the puzzle to help you decide

when they make a reference to a single digit that you cannot otherwise identify as the target or the distractor. Some responses will not give reference to either prevalent digit or to the notion that there are such digits, in which case you would choose the *Neither* option. If there appears to be a vague indication of the prevalent digits, but it is difficult to be certain, you would choose the *Vague or uncertain* option.

2. Awareness of error if incorrect. If the participant's response is incorrect, does the participant indicate that they have detected their error with an explanation of how they know they were wrong? Please answer this question by selecting an option in the **AoE** column whenever the **Actual** column is *Incorrect*. Note that the cell will be shaded gray if the participant correctly solved the puzzle to indicate that you do not need to fill in this cell in that case. Please refer carefully to the solution **Guide** for the detailed criteria for the alternative responses to this question.

3. Basis for choice. Did the participant provide a basis for choosing the response that they gave, (or for their correction, if they noticed and explained their error) and if so, what was it? Please answer this question by filling in the **Basis for Choice** column. We have identified several alternative possible bases for responses which you may find in the solution **Guide tab**. You can find the reference puzzle to these example responses on the last page of this document.

Categories A-C provide a valid basis for making a choice that will always yield the correct answer given the rules of Sudoku and the way the puzzles were constructed. Specifically, in every puzzle, there are always three instances of both the target and the distractor digit, and the target always constrains an empty blue cell that is not constrained by the distractor, so that the distractor, but not the target can be placed in this cell. Noticing this is sufficient to solve our puzzles, given the design constraints. Category D is for responses that may be valid, but are too vague to allow you to be certain. Categories E-G cover responses that would not reliably give the correct answer. Category H is for responses that seem to attempt to provide information about how the choice was made, but are unclear, incorrect, confusing, or fail to specify the procedure used to select between the PDs. Category I is for responses that do not address the question at all.

We believe all the responses can be placed in Categories A-I, but we have added Category J (other) for when you feel that a response does not belong in any of our prespecified categories. If you choose this alternative, use the **Additional Notes** column to describe as best you can what you think the basis was that the participant was using and why none of the provided categories seem to capture this.

You may find certain responses lie near the borderline between two categories. In these cases, please choose the category you feel is the best choice, and use the **Second Choice** column to provide what you think is the next best alternative. For instance, if you find a response that hints at a particular option but is still too vague or confusing to be certain, you may use that option and Category H (in whichever order you think most appropriate) to indicate your uncertainty. Some participants may also have described a strategy that they often used or would have used in a counterfactual setting. Please use the **Second Choice** column to indicate other bases that the participants described. When using the **Second Choice** column, please include additional details in the **Additional Notes** column to explain the uncertainty or to note the mention of an alternative basis.

Please read the **Guide** page in the spreadsheet carefully to familiarize yourself with the classification categories for all of the questions we have asked you to answer before you start rating participants' responses. Note the description of each classification category, the examples, and the comments if any about each of these categories. Please refer to the example puzzle at the end of this document when considering the example responses in the guide. In addition, please bear in mind that the participants may be attempting to convey a correct understanding but may not be using clear or consistent terminology and may even make slips, as anyone can do when they speak or write. Thus, if you think the participant merely mis-stated a correct understanding, you should choose the category that captures what you think the participant intended. If you think it is also possible that they were actually not clear in what they meant, then you should also select Category H, as discussed in the preceding paragraph.

Once you start scoring, please proceed through the participants one by one, from the top of the list to the bottom. However, as noted above, you may go back and revise your previous answers. We expect the task to require your careful concentration and that it will take up to one hour to complete the initial rating of 20 participants. We hope that you will recheck your responses carefully to ensure you have minimized any lapses or cross check any cases where you were uncertain, recognizing that we all have occasional lapses of attention and that the judgements required here are based on ambiguous text in many cases.

We recommend that you have the **Guide** and **Ratings** tabs of the spreadsheet simultaneously visible in two separate windows, as well as the corresponding puzzle, along with a print-out of these instructions. Below we show one possible screen layout, though you may arrange it however you wish, or you may choose to print the Guide if that will make it easier to refer to while performing the ratings.

Thank you again for your assistance with this.

Example Layout

The image displays a Google Sheets interface with two overlapping spreadsheets. The top spreadsheet, titled "Questionnaire Ratings", is a data table with columns for Subject ID, Evaluation, Actual, Response, PD, AoE, Basis for Choice, Second Choice, and Additional Notes. The bottom spreadsheet, titled "Questionnaire (Subject 1)", shows a 6x6 grid with numbers and a table of responses.

Questionnaire (Subject 1) Grid:

		5			
			8		
		8			
			1		5 8
	5	7			
	8	3			

Questionnaire (Subject 1) Table:

Type	Description	Example 1	Example 2	Comments
A. Choose PD that cannot go elsewhere	The participant states that one of the prevalent digits cannot go in any of the empty blue cells and gives that digit as the answer.	I noticed that 3 and 6 occurred together in a box, and focused on these digits. I then noticed that the 3 could not go in any of the empty blue cells, so I chose this digit.	I noticed that one of the two digits that occurred frequently could not go in any of the empty blue cells, so I chose that digit.	Example 1 and 2 are both correct answers.
B. One PD can go elsewhere, choose the other	As a shortcut variant of category A, participant states that they found a cell in the target house that was only constrained by one of the two prevalent digits and chose that digit as the answer. To be valid, this must be a cell that is not in fact constrained by the distractor.	Ignoring the box with three blue cells, I looked for an empty blue cell that only had a 3 in its column. I used this to decide that 3 is the correct answer.	I noticed the two digits that occurred frequently, and noticed that one of them could go in one of the empty blue cells, so I chose the other one as my answer.	This is a constraint, but the participant did not mention it.
C. Found cell where one PD could go and the other could not	The participant mentions finding a blue cell that can contain one of the two prevalent digits and not the other, and chooses the digit that cannot go in that cell as the answer.	I noticed that 3 and 6 occurred frequently in the puzzle, and I knew that answer was one of these two digits. I then noticed that a 6 could go in one of the blue cells but the 3 could not go there, so this meant that 3 must	There was a blue cell where one of the two frequent digits could go but the other could not. The correct answer is the digit that cannot go in this cell.	As with constraint category 1, the participant did not mention it.

Example Puzzle

	6							
				6				
2	4		3		8			
							3	6
				3				
		3						