Considering the Person in the Puzzle: Challenging common assumptions about Sudoku player strategies

Alice Lynch, Chris Jefferson
University of St Andrews
Jack Cole Building
North Haugh
St Andrews
KY16 9SX
al254@st-andrews.ac.uk, caj21@st-andrews.ac.uk

Uta Hinrichs
Informatics Forum
10 Crichton Street
Edinburgh
EH8 9AB
uhinrich@ed.ac.uk

ABSTRACT
Pen & paper puzzle games are an extremely popular pastime, often enjoyed by demographics normally not considered to be gamers. There has been extensive research into generating and efficiently solving digital pen and paper puzzle games, often by creating an artificial player. However, there have been few academic studies focusing on players themselves. We conducted a qualitative study where we observed the Sudoku solving strategies of 31 participants. Our findings reveal interesting discrepancies between common assumptions about players' Sudoku solving strategies made by both guides and AI Sudoku systems, and their actual approach. For example, in contrast to approaching Sudokus in a systematic way and applying simple deductions—a strategy commonly assumed by AI systems—we found that a range and combination of strategies are applied to even the simplest Sudokus. Our findings suggest new directions for designers (both human and AI) of Sudoku and other puzzles, informed by players rather than models.

Keywords
Paper puzzle games, puzzle, Sudoku, Casual Games, Empirical studies, players

INTRODUCTION
Pen and paper puzzle games – puzzles that are solvable by a person without the need to guess, and which have a single solution – are a popular form of casual gaming. They are frequently digitized and published via apps, websites and games. Examples of pen and paper puzzle games include Sudoku, Binario, StarBattle and Tents & Trees.

There has been extensive research into solving and generating pen and paper puzzles computationally (Chatterjee et al., 2014; Howell et al., 2017; Hunt et al., 2007; Maji et al., 2016; Martin et al., 2007; Pelánek, 2014; Reeson et al., 2007). However, handmade puzzles are still considered a selling point and most competitions prefer handmade

Proceedings of DiGRA 2020

puzzles. There are many reasons for this, primarily, the aesthetics of the puzzle, a sense of narrative, and the consistency of the challenge presented by the puzzle (Kanamoto, n.d.).

Some puzzle elements that a player finds challenging will be computationally trivial, while others will be trivial to the player but computationally challenging (Terveen, 1995). This naturally has an impact on computationally modelling players. AI models of human behaviour are, by necessity, based on assumptions; frequently the assumptions that players are systematic, rational, consistent and infallible (Gershman et al., 2015; Kenaw, 2008; Lewis et al., 2014; Noti, 2017). The assumptions made by puzzle game AI are often based on solving or design guides which make similar, though softer, assumptions about players (Espasa Arxer et al., 2021; Pelánek, 2014).

Sudoku

Sudoku is a world-wide phenomenon, appearing daily in newspapers around the world. There are hundreds of implementations of Sudoku available for Android phones and thousands of Sudoku books available on Amazon. Sudokus can be created either via computational generation or by hand. However, computationally generated Sudokus are still considered inferior to hand-crafted Sudokus. The World Sudoku Championship uses handmade Sudokus that are tested by human solvers before their use in the competition, and the use of handmade Sudokus is often used in the advertising of books and magazines, all of which suggests that the automatic creation and grading of Sudoku puzzles could be improved (*27th World Puzzle Championship Instruction Booklet*, 2018; Alexander et al., 2013; Kanamoto, n.d.; Martin et al., 2007).

Better understanding of the process that players use to solve puzzles could help puzzle designers and graders (human or AI). Furthermore, better understanding of the impact that errors and error recovery has on player experience and progress could be used to improve feedback systems and reduce frustration. This understanding would support the design of better systems to scaffold learning processes and support educational efforts that make use of puzzle games (Anderson et al., 1985; Butler et al., 2017; Crute & Myers, 2007; Lee et al., 2014).

In this paper we present a qualitative study examining the behavior of human players. Our findings reveal a range of discrepancies between player's Sudoku strategies and the assumptions underlying even the best Sudoku AI. We hope that these discrepancies can be further explored in future studies and result in both a better understanding of players and improved player experience.

BACKGROUND

Sudoku is a logic puzzle consisting of a 9x9 grid of cells split into nine 3x3 boxes. Each cell in the grid is either empty or contains a digit between 1 and 9, as shown in Figure 1, left. The goal is to complete the grid such that every row, column and 3x3 box contains all of the digits from 1 to 9 exactly once, as shown in Figure 1, right. A valid Sudoku has exactly one solution. Sudokus are traditionally solved using a series of logical deductions.

Definitions

Throughout this paper we use the following terminology:

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

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

Figure 1 Example Sudoku puzzle (left) with solution (right), green highlighting indicates the digits filled in by the player

Candidates: digits which could be placed in a particular cell.

Annotation: any notes, anywhere on the page, that assist the player in storing information about the state of the puzzle but which are not the final digit choices.

Candidate filling: making an annotation inside a cell giving a candidate for that cell.

Completing: writing the chosen digit in a cell.

Dimension: an individual row, column or 3x3 box (as indicated by the bold lines in the puzzle).

Overlapping Dimensions: the combination of the row, column and 3x3 box that overlap a given cell, as shown in Figure 2.

rXcY: to indicate a particular cell. X indicates the row and Y indicates the column. Numbering starts at in the top left, so the top left cell is r1c1 and the bottom right cell is r9c9.

Clue: the digits present in the initial puzzle state.

Minimum Solve: a list of the simplest techniques required to solve a specific Sudoku, chosen from a list of Sudoku techniques ordered by a difficulty measure. It may not indicate the smallest possible number of steps required to solve the puzzle, as the puzzle may be solvable in fewer steps using more challenging techniques (A. Stuart, 2008).

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

Figure 2: Example of Overlapping Dimensions. Cell is marked in black, with the three overlapping dimensions highlighted in grey.

Solving Sudokus

Due to the popularity of Sudokus there are many books, online resources and discussions of solving approaches, many of which include a hierarchy of techniques ordered by difficulty. The following is a summary of the techniques which we refer to in this paper, given in Stuart's order of difficulty (2007). We do not attempt to capture all the different terms and variations which occur in the literature.

Basic Techniques.

The most basic techniques use digits already present in the overlapping dimensions of a given cell. There are two approaches:

Naked Single: If eight of the nine possible digits are already present in the overlapping dimensions of a cell, this cell must take the remaining digit. For example, the top right cell of Figure 2 must be 2.

Hidden Single: This technique looks at a single dimension and finds a digit which can only occur in a single cell of this dimension. For example, the black cell in Figure 2 is the only place in the column where 8 can go.

Many Sudokus can be solved using only these techniques.

Subset-Based Techniques.

The goal of advanced techniques is to eliminate candidates from cells until a Naked or Hidden Single presents itself. It is assumed that players using advanced techniques begin by systematic candidate filling (A. Stuart, 2008).

Naked Pair: Considered the simplest technique after Naked/Hidden Singles. A Naked Pair occurs when the only candidates for two cells in the same dimension are the same two digits and nothing else. This implies these digits must occur in these two cells (in some order) and therefore they can be eliminated from every other cell in the dimension.

Hidden Pair: This technique is the inverse of Naked Pair and is considered a greater challenge. A Hidden Pair occurs when there are two digits which only occur as candidates in two cells of a dimension. This implies these digits must occur in these two cells and therefore all other candidates can be eliminated from these two cells.

Naked and Hidden Pairs can be extended to Naked/Hidden Triples, which occur with 3 candidates and 3 cells. These are assumed to be more challenging, although there is no consensus on whether a Hidden Pair is easier or harder than a Naked Triple.

Pointing Pairs and Box/Line reduction: Pointing Pairs (and Box Line reductions) occur when the only places that a digit appears in one dimension overlaps with a second dimension. For example, consider a 3x3 box where the only places that 2 is a candidate are in the top-most row. The digit 2 can be removed from all other cells in that row, as we know the 2 of that row must occur in this box.

Cycles

The Naked/Hidden Pair approach has been developed into a family of techniques known as ‘fishy cycles’, which we outline here. A full explanation can be found in (Rosenhouse & Taalman, 2012).

X-Wing: The simplest fishy-cycle, an X-Wing consists of 4 cells, with at least one candidate in common, which form the corners of a square. An example is given in Figure 3.

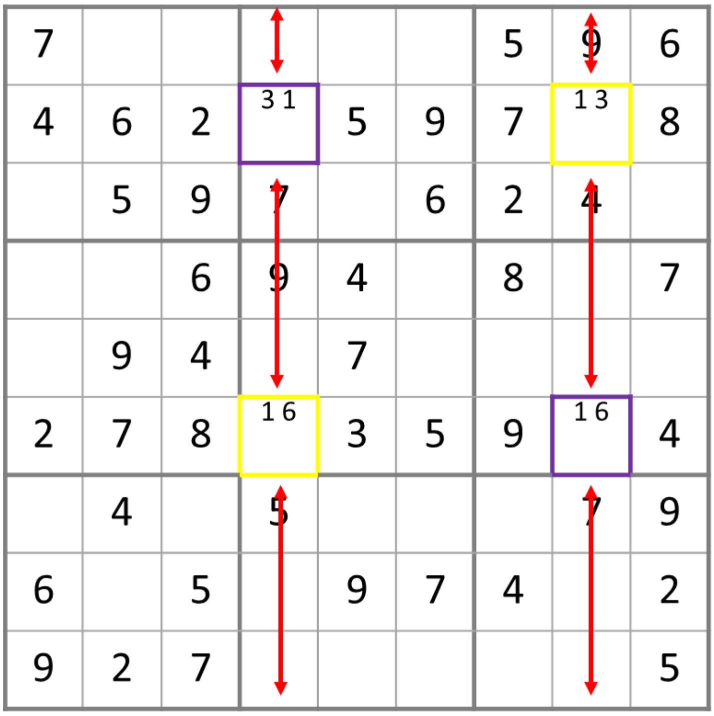

Figure 3: The 4 highlighted squares form the X-wing and 1 can be excluded from the columns indicated by red arrows

The larger cycles work on the same pattern as the X-Wing but involve a greater number of cells.

Esoteric Techniques

There are a variety of more advanced techniques, including unique rectangle, many of which are quite rare (A Stuart 2007; A. C. Stuart n.d.).

Chains

Chain techniques are assumed to be one of the last resort options. The player guesses a digit and then follows the chain of deductions until they either complete the Sudoku, run out of deductions or reach a contradiction that can be used to eliminate all candidates involved in the chain. Players often use colors to track the chain. There are several methods of performing “Chain” reasoning (A. Stuart, 2008).

STUDYING PLAYER’S APPROACHES TO SUDOKU SOLVING

We designed a qualitative study to investigate how players’ solve Sudokus and to explore differences between these processes and assumptions made by guides and AI models. The study design was informed by an online survey, used to gather initial information about Sudoku players’ typical practices and preferences when solving Sudokus.

Preliminary Online Survey

This was conducted to understand trends in the Sudoku community, it focused on the medium used to play Sudoku, and whether solving techniques were researched. It was distributed via university mailing lists and social media; 522 participants completed it. Here we only discuss the findings that informed the main study design.

The main findings that informed the study design were that, as shown in Table 1, 391 of the 522 respondents solve Sudokus using paper and pen. 122 of respondents (24%) reported that they researched solving techniques.

Type of Tool	Number of Participants
Digital	115
Paper	204
Digital and Paper	187
Other	1

Table 1: Summary of type of medium used by participants when solving Sudoku.

The main design decision resulting from the survey was that the study should be conducted on paper, as the majority of respondents were familiar with it, and it avoided the introduction of assumptions about the notations a player might use.

Motivation for the In-Person Study

Our literature review highlighted that current guides and computational models have mostly been validated against each other and/or the solve time rather than the player, yet they are based on assumptions about the player (Hunt et al., 2007; Pelánek, 2014). The most sophisticated computational models all rely on two key assumptions: that at every step players randomly choose one of the easiest available moves, and that the definition of the easiest move is consistent amongst players. However, our survey demonstrated that most players do not research Sudoku solving techniques and therefore, the strategies that they invent, may not match the published guides. Their perception of the easiest move will be based on their own experience and may not reflect the "correct" approach as defined by published guides. Particularly as the order of difficulty is not consistent between the guides (A. Stuart, 2008; *Sudoku Dragon: Sudoku Puzzle Solving Strategies*, n.d.; *Sudoku Snake - Solving Techniques*, n.d.).

To understand how robust the underlying assumptions are we explore how players approach and experience solving Sudoku, we designed an exploratory qualitative study with the following question in mind:

How accurate are the following, foundational assumptions of the current best models of Sudoku players?

1. Players solve puzzles using the named techniques described in the literature.
2. The only annotation which players use is listing the allowable candidates for each cell.
3. Players solve puzzles by repeatedly choosing a move of easiest difficulty.
4. Players do not make mistakes while solving puzzles.

Participants

Participants for our study were recruited from the pool of participants who took part in the online survey (28) and university members that expressed interest and then completed the online survey (3).

We had more than 31 responses and chose respondents that, when combined, provided the best range and distribution of expertise levels. Unfortunately, very few respondents were experts or novices. 29 participants had previous solving experience, 6 had only used digital tools. All participants participated in person and were compensated with a £10 local book token.

We recruited 31 participants (20 female; 11 male) with a median age bracket of 25-34 (Figure 4). The educational background of our participants is not representative, 27/31 had finished an undergraduate degree. We did not ask participants if they were a member of the university.

Expertise Rating	Number of Participants
Complete Novice	2
Beginner	7
Intermediate	17
Advanced	5

Table 2: Participants' self-ratings of Sudoku expertise

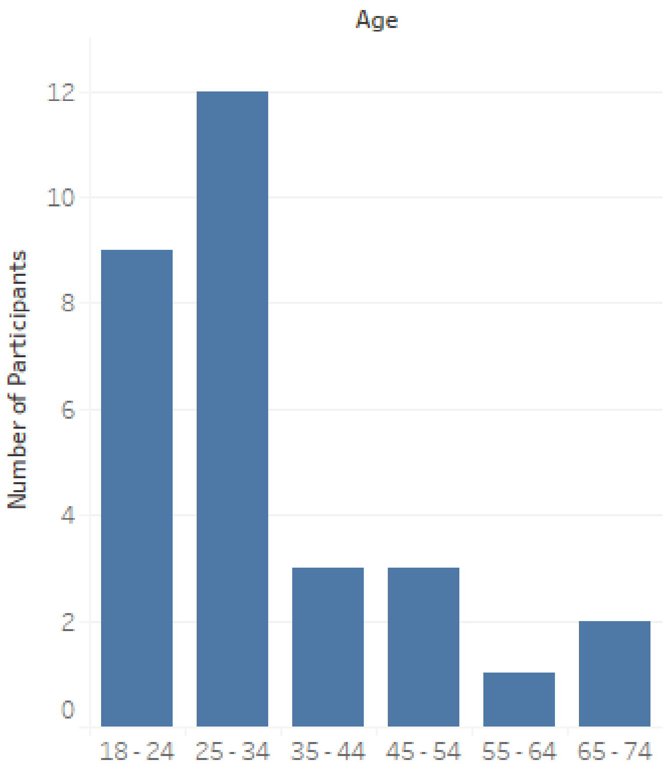

Figure 4: Age distribution of participants

Study Design & Procedure

The study took place in a private office where participants solved Sudokus. Participants had individual sessions, they did not overlap with any other participant. Each study session (both 1 & 2) started with a pre-study questionnaire. The Session 1 questionnaire asked participants to rate their expertise (Table 2), their experience at solving Sudokus, how recently they had solved a Sudoku and whether they had prepared for the study. The Session 2 questionnaire was similar: they were asked to rate their expertise (in case they had reconsidered), how many Sudokus they had solved since the first session and how recently they had solved a Sudoku. They were also asked if they had researched Sudoku since Session 1.

In both sessions, participants were then given the first puzzle (and a duplicate in case they wanted to restart) to solve. We explained and provided printed Sudoku rules.

Participants were left alone, when the participant finished or wished to stop they alerted the experimenter, who waited out of view outside. They were asked to fill in a questionnaire which asked them to describe (or name) the solving techniques they had used, to rate how challenging, enjoyable, and frustrating they found the Sudoku (using a 7-point Likert scale where 1 indicates none and 7 indicates extreme), and any further comments. They were then given another Sudoku.

In both sessions, participants were given Sudokus for the duration of the allotted hour, at which point participants were given the option to continue the current puzzle, after 1.5 hours the study session was ended. Once they finished, a short interview was conducted by the researcher. This approach was chosen over a think aloud study due to concerns about the additional cognitive load interfering with the solving process, particularly as the impact on working memory might bias participants to a more note based approach than they would otherwise employ (Branch, 2000; Hoppmann, 2009; Van Someren et al., 1994).

The researcher, during the interviews, asked participants to talk through their approach to solving the puzzles and which they enjoyed or found frustrating. They were also asked to explain any notation they used and if they came up with any new approaches.

We did not impose a time limit on individual Sudokus. Sudokus were provided to participants on sheets of A4 paper. Participants were provided with plain and coloured pens and pencils, a rubber and a pencil sharpener.

Session 1

This session's goal was to gain insights into participants' overarching approach to solving Sudoku and what challenges they encountered. Participants were asked to completely solve the provided Sudokus. The Sudokus participants were given were selected based on their rating of the the previous Sudoku solved. This was done to accommodate the range of expertise. Beginner players would be unable to solve Sudokus requiring advanced techniques; which would not provide useful information and would frustrate participants.

Sudoku Design: Session 1 required Sudokus of varying difficulties which could be solved fast enough that participants could solve several during the session. The Sudokus were created by the authors, in some cases based on existing Sudoku from SudokuWiki (A. Stuart, 2008). To facilitate this, each Sudoku was tested, first using the online solver at SudokuWiki (A. Stuart, 2008), to establish the minimum solve under Stuart's difficulty ordering (A. Stuart, 2007). The expected time to solve the Sudokus was tested by the authors. 14 Sudokus were selected for Session 1 (Table 5), categorised into 4 difficulty classes (Table 3).

Difficulty Class	**Techniques Required for minimum solve**
Very Easy	Naked singles
Easy	Hidden singles
Medium	Simple Subset techniques
Hard	Any further techniques.

Table 3: Sudoku Difficulty classes, based on expected difficulty as discussed in the Solving Sudokus Section. Each class includes all techniques used by previous classes.

ID	**Level**	**Empty Cells**	**Technique Required (Min. Solve)**
R	Very Easy	23	Naked Singles
D	Very Easy	36	Naked Singles
Q	Very Easy	19	Naked Singles

K	Very Easy	28	Naked Singles
F	Very Easy	41	Naked Singles
V	Easy	24	Hidden and Naked Singles
Y	Easy	24	Hidden and Naked Singles
All Sudokus below include Hidden and Naked Singles			
M	Medium	27	1 Naked Pair
X	Medium	27	2 Naked Pairs
T	Medium	32	1 Naked Triple
S	Medium	38	1 Naked Triple
Z	Hard	32	1 X-Wing, 1 Naked Pair
L	Hard	22	1 Simple Colouring, 2 Naked Pairs
AA	Hard	33	1 Simple Colouring, 1 Swordfish

Table 4: Puzzles included in Session 1, with a difficulty class based on existing literature, the total empty cells, and the techniques required for a minimum solve.

Session 2

This session's goal was to examine what impact particular techniques had on participants' enjoyment, frustration and challenge, and whether the expected technique was even used. The Sudokus were designed to require a particular technique, based on the Stuart difficulty ordering (A. Stuart, 2008). Participants were asked to either fill in the Sudoku until they had filled in a highlighted cell (CPC) or until they had completed X cells (CXC).

This allowed more puzzles to be completed in the available time. The techniques were not described or named and the participants were not aware which technique was being tested by each puzzle. Participants had the puzzles presented to them one at a time, in the order in Table 5.

Sudoku Design: The final chosen Sudokus chosen are listed in Table 5. For puzzles E, P, & W which required systematic candidate filling we gave participants an extra copy with possible candidates filled in, but without all impossible candidates excluded. This aimed to help participants by reducing mechanical effort.

ID	Technique Being Tested
A	Complete a row/column missing with one empty square, then Naked Single. (CPC)
B	Naked Single, then another Naked Single (CPC)
U	Two Hidden Singles (CXC)
C	Naked Pair, then Hidden Single (CXC)
O	Hidden Pair, then Hidden Single (CPC)
E	X-wing, then Naked Single (CXC)
P	Unique Rectangle, then Hidden Single (CXC)
W	Simple colouring (CXC)

Table 5: Puzzle types included in Session 2, with IDs, in the order they were presented to participants.

Data Collection and Analysis

Video-audio data captured participants activity (see Figure 5).

Figure 5: Primary (Left) and Secondary (Right) Camera Angle

Participants received paper questionnaires which were transcribed. Interviews were conducted after each session and recorded.

We conducted a broad-stroke analysis of all participants' video. We noted the time taken per puzzle, the order in which cells were filled in and any annotations or mistakes. We used an open coding approach, where the coding scheme was iteratively developed during the analysis and then finalised amongst the researchers (Saldaña, 2015). The questionnaire results were plotted, descriptive statistics were generated and the textual responses were coded.

We then selected 9 participants to analyse in greater detail. These participants were chosen to provide a broad representative sample of notation styles and approaches. 3 participants from each self-selected competence level were analysed following accepted qualitative analysis practice (Heath et al., 2010; Marshall & Rossman, 2014; Saldaña, 2015). This focused on the finer details of a participants' process: the type of annotation used, techniques employed while solving and what mistakes they made and why.

FINDINGS

In this section we discuss our observations of participants' interactions with the study tasks. We first focus on the participants' process when solving Sudokus, then look at the types of errors made by participants. Finally, we discuss the participants' perception of challenge. In the next Section we will discuss the implications of these findings.

Processes of Solving Sudokus

We found that annotation approaches varied widely between participants. While many participants noted down potential candidates in some cells, only 1/9 of the participants we coded in detail were mostly systematic about writing down all potential candidates in every empty cell. No participant systematically removed candidates at every stage as they progressed through the puzzle and started completing cells.

Notation

We found that the majority of participants (29/31) made some sort of annotation on at least one puzzle, as shown in Figure 11; this trend was inverted for puzzle U where very few participants used annotation. We observed a range of approaches to annotation and categorised them into the following:

Systematic Candidate Filling: This is the approach commonly described in the literature, systematically filling in all potential candidates in all cells in the grid. This approach was rarely employed by participants. Interestingly, 5 participants believed they were using this approach despite it not being consistently observed in their solving.

Local Candidate filling: An approach where all possible candidates in all the cells in a particular row/column/box were filled in, as shown in Fig 6. This was rarely described by participants but was frequently observed in the video data.

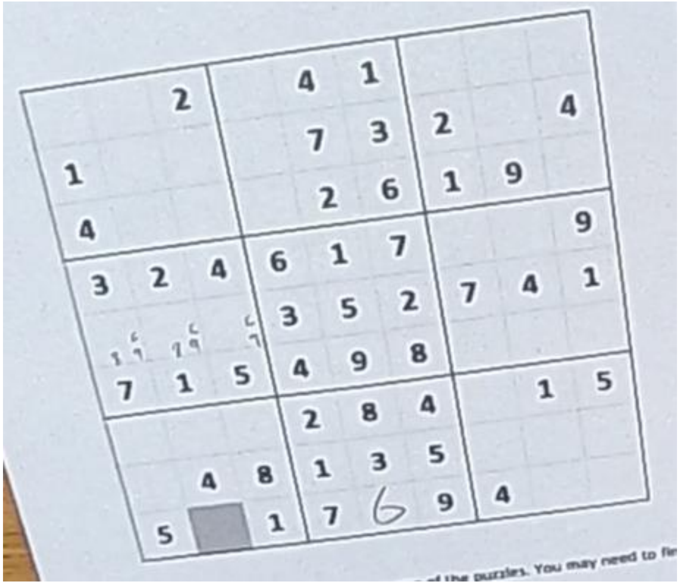

Figure 6: Example of local filling, the participant only filled in candidates for cells in row 5

As & When Digit filling: An approach where the participant made notes about a particular digit, without completing all potential candidates in a cell. This approach was rarely described but was frequently observed. P10[1] acknowledged this approach, describing it as "...it goes with stream of thoughts, because I don't want to be selective".

Small Set Candidate Filling: An approach where the player only makes a note of a particular candidate's possible positions in a dimension if there are only two possible positions, as shown in Figure 7. 7 participants explicitly described this process, P9 described this approach as "you'd have a look to see if there's any way with only two numbers, so then you'd put in what they were, so that's a 1 and 5, so you'd write in a little one and a little five". P32[2] clarified that they preferred this notation to filling in all the possibilities as "otherwise it gets very messy and just confuses me".

---

[1] All participants are referred to as P followed by their participant number.

[2] Participant IDs start at 7, 1-6 were used in the initial development stages

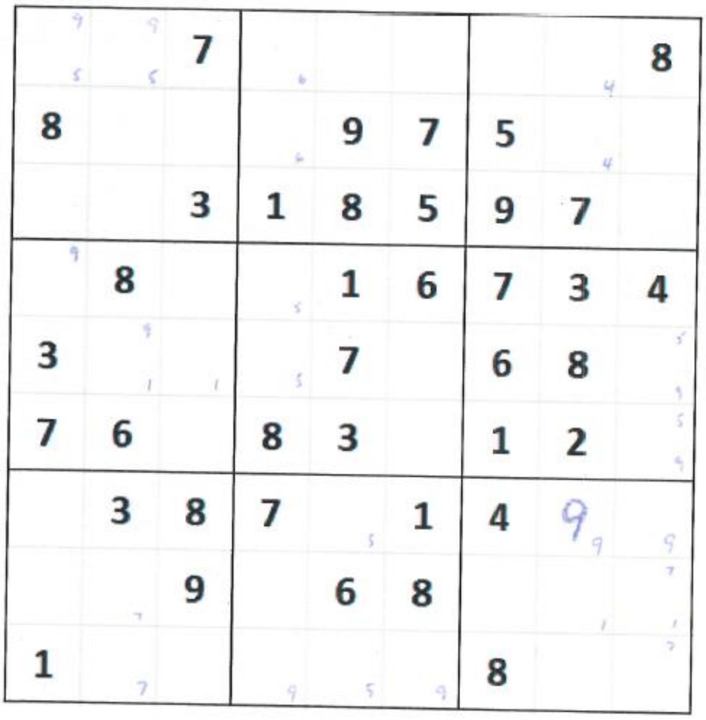

Figure 7: Example of participant using Small Set Notation - they have only noted candidates when they appeared twice in a dimension

Dimension Candidate Notation: An approach where the player notes all candidates that can be put (anywhere) in a dimension, as shown in Figure 8. P16 described their process as “I go through each rule, normally bottom to top and then I’ll go, is there a one there [in the row] and write it at the side”.

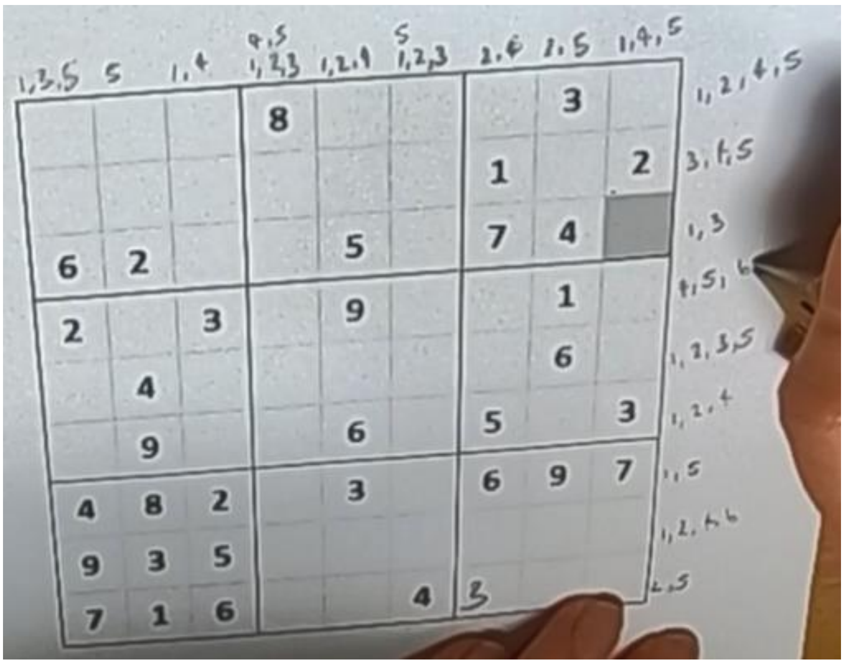

Figure 8: Example of Dimension Candidate Notation. The participant is writing the candidates missing from each row and column, noting all occurrences of each digit in order

Excluded Dimension Candidate Notation: The player notes all the digits already present in the dimension.

Highlighted Cell focused annotation: Shown in Figure 9, this phenomena is a result of the experiment design and not generalizable.

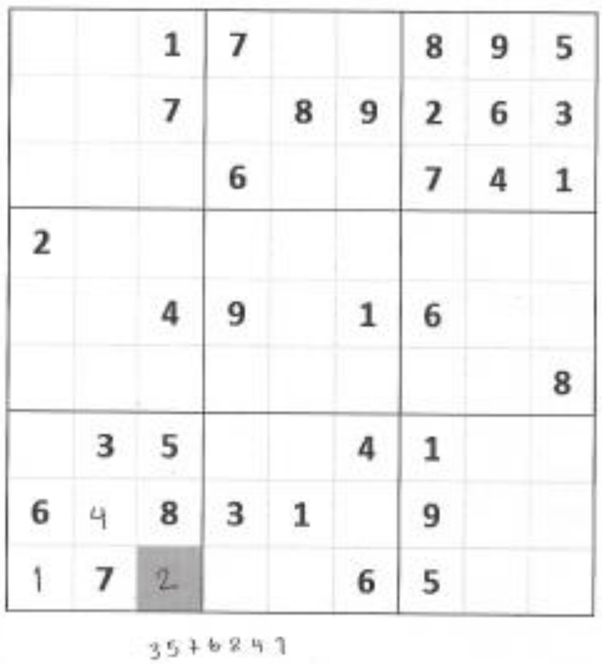

Figure 9: Example of Highlighted Cell notation

Other Approaches: Some participants used unique alternatives or extensions to the above which weren't categorized (an example is shown in Figure 10).

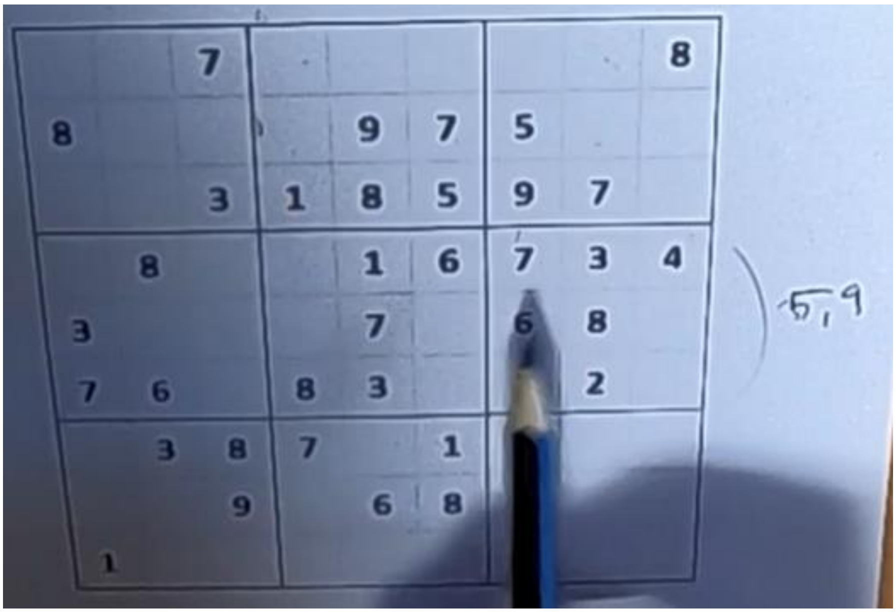

Figure 10: Unusual notation used by one participant to indicate Naked Pairs.

Some participants eventually filled in all cells, but filled in clumps, completing a single dimension or cell at a time. These participants tried to complete cells in between flurries of candidate filling, often failing to remove candidates rendered impossible by the completed digits. Most participants filled in candidates in some sections and ignored them elsewhere. Participants used different annotations on different puzzles. The need for annotation seemed to impact participants' solving experience, P14 stated "I find it less fun if I have to write down the smaller numbers".

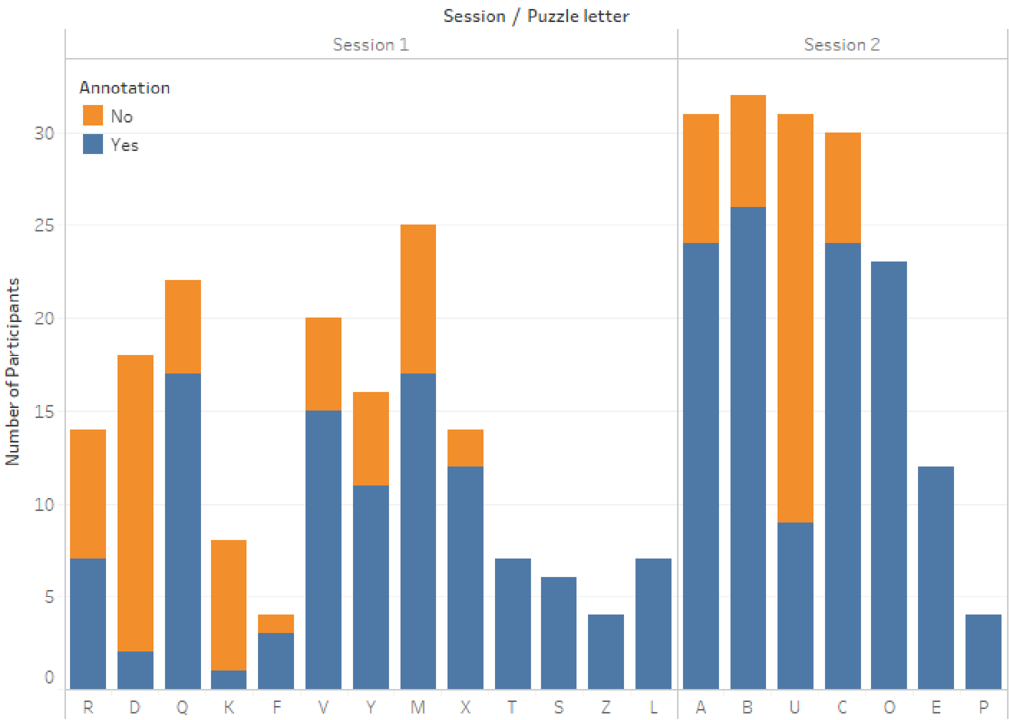

Figure 11: Use of annotation by participants (The B, U, and C puzzles had at least one participant do two puzzles, annotating only one)

P18 was the only participant to never use annotations. They successfully solved puzzles requiring naked pairs in the first session, though they considered puzzle C to be impossible to solve without a 'leap of faith'. Some other participants managed to solve naked pair puzzles without annotation - notably 4/6 participants who attempted C without annotation were successful.

Order of Play

The order in which players solve a Sudoku is expected to be somewhat pre-determined by the puzzle design. However, participants frequently started in cells not considered entry points and followed varied routes through the puzzle (see Figure 13). This was particularly noticeable in the M puzzles, which were expected to have a single entry point, based on Stuart's solver (2008; 2007). Of the 21 participants that attempted M1 (including 8 who made mistakes), only a single participant started with the expected cell. The paths participants took can be seen in Figure 13. The entry via A9 and C9 suggests participants did a pointing pair before the expected naked pair (which would have removed the pointing pair).

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

Figure 12: Puzzle Q with entry points highlighted

The first column of R1 contained only one empty cell, so we expected it to be the primary entry point. However, only 1/17 of the participants that attempted R1 (including 4 that made mistakes) completed this cell first. A numerical approach may have influenced this, as this cell took 3 while others required 2. In contrast, Q1 had 3 expected entry points (see Figure 12), however 12/13 participants that attempted it started in the same cell (r7c2).

The interviews and questionnaires provided insight into the participants' approach to the puzzles. Some participants described looking for easy approaches and then applying harder ones. "I try the easy tactics and if they don't work I have to try some more difficult tactics" - P7. However, 9/31 looked for dimensions with the fewest empty cells, P15 specifically noted that they look for the most filled 3x3 box before looking at rows and columns. 17/21 explained that they approach the Sudoku in numerical order - either checking all the ones, then twos etc until the nines or in reverse. P34 described "I'd try and do all the ones and then try and do all the twos and then try and do all the threes etc". Those that worked through easiest to hardest techniques described different difficulty orderings.

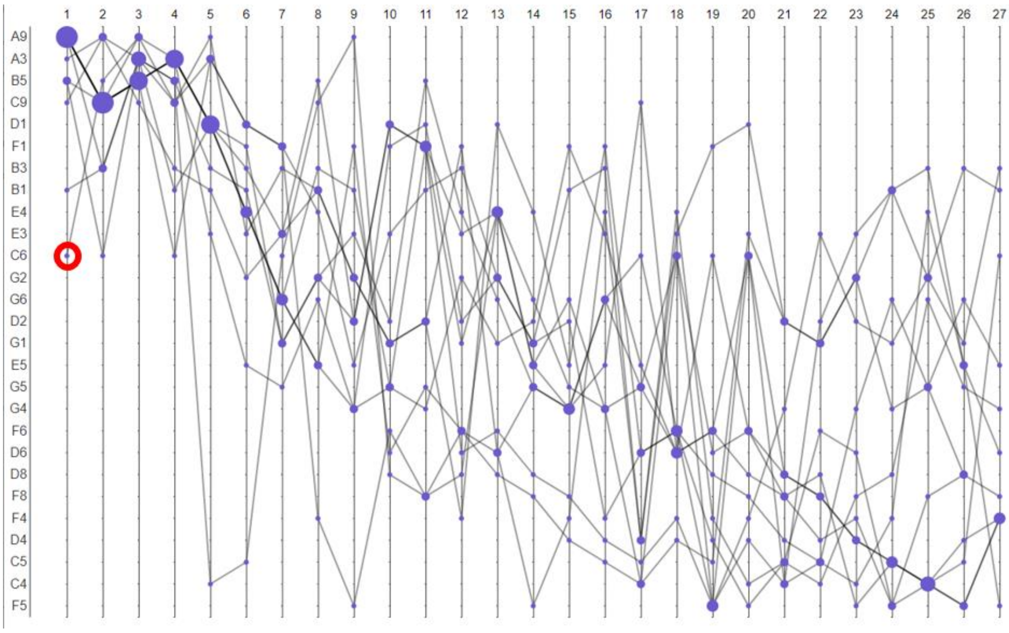

Figure 13: Path taken by participants through M1 puzzle. Circles size: the number of players that filled in that cell at that step. Y-axis: the cell filled in (ordered by the average step they were filled), X-axis: the step. Excludes participants that made a mistake. The expected entry point is circled in red.

Mistakes

20/31 participants made at least one mistake during Session 1, the types of mistakes made are discussed below.

Digit already present: A common error occurred when participants completed a digit in a dimension that already contained that digit. In some cases participants completed digits directly adjacent to the same digit, as shown in Figure 14.

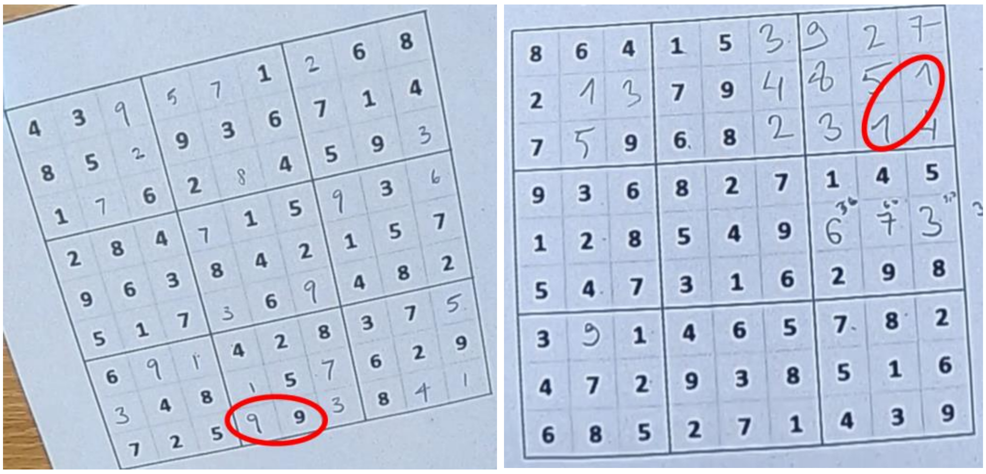

Figure 14: Left: Example of a 9 incorrectly completed in puzzle R1, adjacent to the clue that excluded it, Right: Example of two 1s incorrectly completed in the same box in puzzle Q1.

Incorrect Candidate exclusion: This occurred when a participant incorrectly excluded a candidate from a cell, resulting in an invalid chain of deduction. In the example in Figure 15, the participant incorrectly excluded the digit 8 from the highlighted cell, leaving 9 as the only candidate. The completed digit should have been 8.

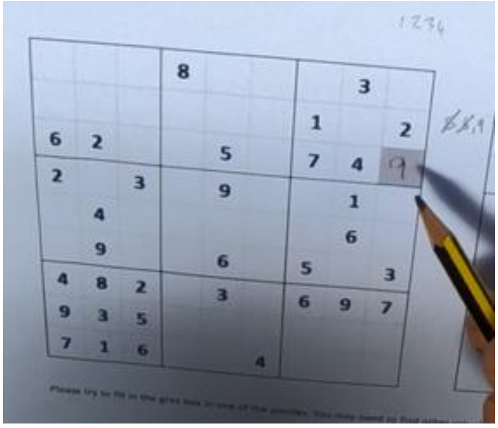

Figure 15: Example of incorrect candidate exclusion in puzzle B2, r3c9 should be 8 not 9

Incorrect Guess: 19 participants stated they guessed when they could not make any more deductions. For example, in the right hand side of Figure 16, the participant stated they guessed (incorrectly) between the two allowed values (2 and 3) for r9c1.

Error propagation: When participants made errors, if they didn't notice them immediately the error propagated through the puzzle. For example, in Figure 16, left, the first error was entering 6 in cell r8c3. This leaves r9c3 as the only place in the bottom left box where 5 can be completed. This is also incorrect, despite the deduction leading to it being correct. Further, this leaves the cell r7c9 as the only valid cell for 5 in the bottom right box. Five participants explicitly commented that the propagation of the errors caused frustration as it was hard to backtrack to the source of an error and they often simply started again.

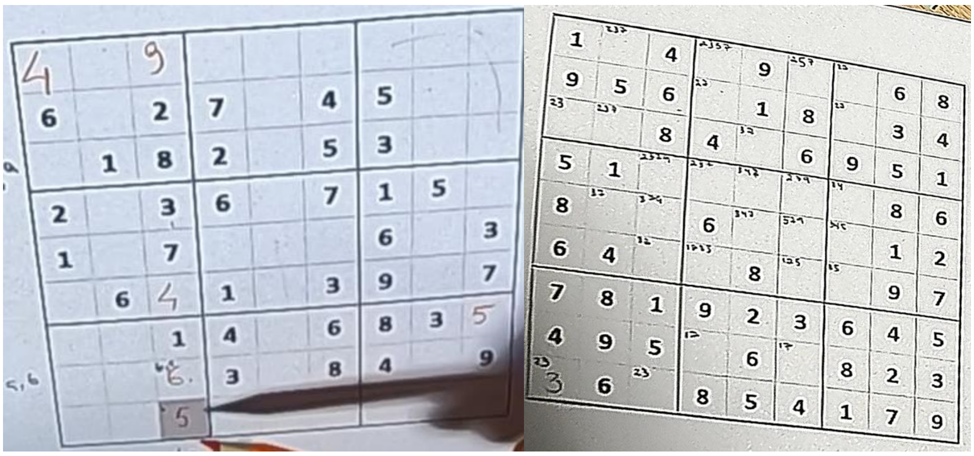

Figure 16: Left: Example of error propagation in puzzle O1, a mistake in excluding 5 from cell r8c3 leads to errors in r9c3 and r7c9, Right: Example of an incorrect guess (as described by participant) in r9c1

Perception of Difficulty

Q and F were found to be harder than the other Very Easy puzzles. While F has the most empty cells, Q had the least. Participants stated they found having one 3x3 box (see Figure 12) in Q completely empty increased the challenge. "The empty box in the middle made it trickier" – P31 "…a particularly snarly one, like when there's one box that's completely empty" - P15. Also regarding Q: "Took a minute to identify where to start as so much was already completed" (P16) and "I prefer an emptier grid to start with and having to fill in more" - P28. P30 however mentioned that they liked having an empty 3x3 box.

Many techniques which participants described were recognizable from the literature, however the difficulty they associated with them varied. Many participants described hidden singles but didn't consider naked singles, or described them as the subsequent step. P14 described hidden singles as 'any obvious numbers'. The difficulty of naked pairs varied between participants, ranging from equal to the singles, to more challenging than hidden and pointing pairs. P11 used them extremely readily: "I kind of use it so much I don't think of it as technique" and "I do that without realizing I'm doing it. I don't think". Some participants found pointing pairs the easiest of the advanced techniques: "It's the one that's easiest for me to approach" – P15.

The number of entry points impacted on perceived difficulty and enjoyment, P15 stated "...if you can fill in a couple of firsts at the beginning and then get stuck, [false start], one of them was like oh, there is more multiple choices, I have to think about it now, I quite enjoy that. Whereas, if I can't fill one out from the beginning it's just a little bit frustrating".

The variation in missing digits also impacted both the perception of difficulty and enjoyment. 3 participants explicitly mentioned that they found puzzles where all the instances of a digit were missing both harder and less enjoyable, "I'd rather there was more numbers missing but more variety of numbers" – P16.

Player's challenge ratings indicated that, in Session 2, U was considered least challenging, despite requiring Hidden Singles instead of Naked Singles. This is supported by the time taken to complete the puzzle, where U had the lowest average solve time. This was further supported by the interviews, where 5 participants described hidden singles as the easiest technique. Although, 5 other participants described approaching the puzzles by checking all the possible values for each cell without necessarily noting them down and felt this was the easiest technique. There was a notable increase in perceived challenge and time from the puzzles A, B & U (Naked and Hidden Singles) to puzzle C which includes a Naked Pair.

Intermediate and advanced participants rated the Naked Pair as more challenging than the Hidden Pair, while beginner participants rated it less challenging. (Although due to the small sample of beginner participants this may not be representative.) This remained true when participants who didn't attempt the Hidden Pair were excluded from analysis.

Participants often did not follow the expected order of play. The most notable case were participants readily using chain methods in puzzle C which had a minimum solve of a Naked Pair and participants employing Pointing Pairs in puzzle M, when the minimum solve expected a Naked Pair.

DISCUSSION

Our findings allow us to validate the assumptions underlying all existing computational models. In particular we consider candidate filling, order of deductions, experience of difficulty and the impact of error. In each case we demonstrate evidence that the existing computational models do not align with players and discuss how improved guides, designs and models could be produced.

Notation

Most guides assume players systematically fill in all possible candidates for each cell, then systematically eliminate candidates as solving progresses. Computational models of Sudoku players share this assumption. We instead found extensive variation in the notes players take, and how players use those notes as they solve the puzzle.

Most surprisingly, the systematic approach discussed above rarely occurred in practice. The approaches used by many participants meant that some techniques, such as naked singles, were harder to employ. This suggests future guides and computational models should avoid the assumption that players will behave systematically and consistently.

In particular, many online systems assume players are performing full candidate elimination and mark other notations as ``wrong'', which may confuse or upset players using an alternative notation scheme.

Relative Difficulty of Techniques

Based on existing assumptions we expected:

- a linear increase in difficulty between different techniques
- techniques would be attempted in the established order of difficulty.
- the order in which techniques were applied would have no impact on perceived challenge.

We found all these assumptions to be flawed.

The accepted order of the difficulty of techniques did not apply, as participants found Hidden Singles easier than Naked Singles, this may be related to annotations – finding Naked Singles is easier when using standard candidate elimination. Similarly participants found Hidden Pairs easier than Naked Pairs, this may also be related to non-standard notations or may indicate that it is easier to independently discover the Hidden Pair. 3 participants described pointing pairs and classed them as on par or easier than other subset techniques. One of them, P15, classed them as the easiest technique.

The substantial jump in challenge from Hidden/Naked Singles to Naked/Hidden Pairs suggests that the impact more advanced techniques have on difficulty is greater than previously assumed. This suggests that Sudokus on the cusp of higher difficulty levels should provide players with support in solving the more advanced techniques (for example, by including several examples, all of which can be used to solve the puzzle).

The ready use of chain techniques by participants suggests they may be more intuitive than previously accepted.

It is clear that existing assumptions about techniques' relative challenge and their impact on a puzzle's difficulty are flawed. The finding that participants did not apply techniques in the established order of difficulty, suggests that the ordering needs to be re-evaluated. Players also disagreed on the difficulty of techniques, partially based on their notation approach. This suggests digital tools could learn an individual's perception and adapt accordingly.

Overall, it is clear that more factors need to be considered when assessing difficulty. These findings have implications for teaching and help systems - for example, when guiding players towards the next easiest step. Help systems could also consider that players may resort to chain techniques unnecessarily and provide guidance if they detect a chain technique being used where an alternative is possible.

Order of Play

Figure 11 demonstrates that players can consistently perform a deduction which differs from that suggested by current models. This affects both help systems and difficulty measurement. This suggests a data-driven approach, where the moves taken by a majority of players are used for difficulty modelling and suggested to future players. A full analysis using a data-driven approach would require a much larger data set, therefore it is left for future work.

Spatial layout

The findings that Q and F were considered more challenging than the other very easy puzzles was unexpected. We conclude that the spatial arrangement of the puzzle can impact the challenge a player experiences and therefore should be considered by designers, whether human or AI.

Impact of Error

Frequency and recovery from error contributes to a player's experience in most games and is normally an important consideration for designers. However, in games like Sudoku, where each move should result from a logical deduction, player error is rarely considered. Our findings clearly show that participants made errors, which were often missed -- Figure 14 shows two final submissions with adjacent repeated digits.

Looking at participants' errors, they appear to come from both flawed visual searches, and flawed logical deductions. The common error of completing a digit already present in one of the dimensions overlapping the cell often results from a flawed visual search. This error occurred even when the participant had completed the other occurrence of the digit themselves, strongly implying that players have a limited ability to store the state of the puzzle in their head.

Errors once made, propagate, as the grid is now in an incorrect state, so information deduced from it is flawed. Error propagation is a key contributing factor in making error recovery challenging. Participants reported increased frustration when they made an error, as it often resulted in them restarting, due to recovery being too difficult.

Overall, our findings demonstrate that player error is not unusual and raises questions about the impact mistakes have on players' experience and how puzzle, or interface design could be used to mitigate it. Players will sometimes correct mistakes almost immediately, therefore immediately correcting errors as soon as they appear would reduce player agency. However, unnoticed errors reduce player enjoyment, as they are often unrecoverable without restarting. Better systems to "fix" errors may improve player's enjoyment. For example, allowing the deductions that propagated from the error be tracked and reverted would allow error recovery without restarting the puzzle or losing valid deductions made alongside erroneous ones.

LIMITATIONS

This is a small-scale exploratory study, designed to investigate the assumptions in existing player models. While it demonstrates many previously published assumptions should be reconsidered, it may not be sufficient to provide in-depth guidance to future model designs.

The participants may not be representative of the general population. While this study provides rich qualitative information, it is a small-scale study and some areas would benefit from further large-scale quantitative studies. We leave this to future work.

CONCLUSION

Many of the assumptions existing guides, designers and AI models make about Sudoku players are flawed. The extensive variation, both in notation and logical approaches, strongly suggests that Sudoku design or models based on rigid assumptions regarding player approaches are unlikely to produce puzzles of predictable challenge and reward. Designers (whether human or AI) should consider the different approaches players use when solving Sudokus, including players' different methods of annotation, logical deductions and mistakes. It is also important to explore the narrative that players could take through the puzzle, both the points of challenge and the number of deductions required to complete the puzzle. Treating players as automata who always perform the easiest available technique, and pick randomly if there are several options at the same level, does not reflect players' behaviour. Furthermore, it should be considered that recent steps impact the player's current deductions and focus. It seems likely that in order to produce rewarding puzzles of a predictable difficulty that the different paths which players may take through the puzzle need to be considered.

Perhaps the most interesting outcome of this work is the implication for tutoring systems and scaffolding. Puzzle games are used extensively throughout education, therefore providing better support systems has the potential to increase student attention and engagement. However, it is clear from this research that tutoring systems would need a way to interpret the notation style used, as assuming that a student is systematically noting down all possible options has been shown to be flawed. They may be noting down subsets, or noting down values that have been excluded from possibility.

FUTURE WORK

There are several avenues of research suggested by this study, a more in depth examination of the psychology underlying player's different approaches, possibly large-scale, would provide more information to puzzle generator and interface designers. This study indicates that the way players approach puzzles is more complex and varied than captured in existing player models.

Given the importance of games, both commercially and in education, a more in-depth examination of player motivations and causes of player variance, over a wider range of participants and games would be beneficial.

Producing models for individual players, by observing how play style evolves over time, would provide a significant improvement over existing fixed and general models.

Finally, a similar, open, qualitative study with a digital puzzle interface to examine whether the tool system's assumptions changes player behaviour.

REFERENCES

27th World Puzzle Championship Instruction booklet. (2018). World Sudoku & Puzzle Championship 2018. http://wscwpc2018.cz/wp-content/uploads/2018/10/WPC_2018_instruction_booklet_v3.pdf

Alexander, J. T., Sear, J., & Oikonomou, A. (2013). An investigation of the effects of game difficulty on player enjoyment. *Entertainment Computing*. https://doi.org/10.1016/j.entcom.2012.09.001

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). *Intelligent Tutoring Systems | Science*. *228*(4698), 456–462.

Branch, J. L. (2000). Investigating the information-seeking processes of adolescents: The value of using think alouds and think afters. *Library & Information Science Research*, *22*(4), 371–392.

Butler, Z., Bezakova, I., & Fluet, K. (2017). Pencil Puzzles for Introductory Computer Science: An Experience- and Gender-Neutral Context. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 93–98. https://doi.org/10.1145/3017680.3017765

Chatterjee, S., Paladhi, S., & Chakraborty, R. (2014). A Comparative Study On The Performance CharacteristicsOf Sudoku Solving Algorithms. *IOSR Journal of Computer Engineering*, *16*(5), 69–77. https://doi.org/10.9790/0661-16566977

Crute, T. D., & Myers, S. A. (2007). *Chemistry for Everyone Sudoku Puzzles as Chemistry Learning Tools W* (No. 4). www.cambridgesoft.com/software/ChemDraw/

Espasa Arxer, J., Gent, I. P., Hoffmann, R., Jefferson, C., & Lynch, A. M. (2021). *Using small MUSes to explain how to solve pen and paper puzzles*. https://research-repository.st-andrews.ac.uk/handle/10023/24086

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in Qualitative Research*. SAGE Publications.

Hoppmann, T. K. (2009). Examining the 'point of frustration'. The think-aloud method applied to online search tasks. *Quality & Quantity*, *43*(2), 211–224.

Howell, I., Woodward, R., Choueiry, B. Y., & Bessiere, C. (2017). *Solving Sudoku with Consistency: A Visual and Interactive Approach*. http://sudoku.unl.edu/

Hunt, M., Pong, C., & Tucker, G. (2007). Difficulty-driven sudoku puzzle generation. *The UMAP Journal*, *29*(3), 343–362.

Kanamoto, N. (n.d.). *Why hand made?* Nikoli. Retrieved 13 February 2021, from https://www.nikoli.co.jp/en/puzzles/why_hand_made.html

Kenaw, S. (2008). Hubert L. Dreyfus's Critique of Classical AI and its Rationalist Assumptions. *Minds and Machines*, *18*(2), 227–238. https://doi.org/10.1007/s11023-008-9093-7

Lee, M. J., Bahmani, F., Kwan, I., LaFerte, J., Charters, P., Horvath, A., Luor, F., Cao, J., Law, C., Beswetherick, M., Long, S., Burnett, M., & Ko, A. J. (2014). Principles of a debugging-first puzzle game for computing education. *2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 57–64. https://doi.org/10.1109/VLHCC.2014.6883023

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization. *Topics in Cognitive Science*, *6*(2), 279–311. https://doi.org/10.1111/tops.12086

Maji, A. K., Jana, S., & Pal, R. K. (2016). A Comprehensive Sudoku Instance Generator. In R. Chaki, A. Cortesi, K. Saeed, & N. Chaki (Eds.), *Advanced Computing* *and Systems for Security: Volume 2* (pp. 215–233). Springer India. https://doi.org/10.1007/978-81-322-2653-6_15

Marshall, C., & Rossman, G. (2014). *Designing Qualitative Research—Catherine Marshall, Gretchen B. Rossman—Google Books* (6th ed.). SAGE Publications.

Martin, D., Cross, E., & Alexander, M. (2007). *Cracking the Sudoku: A Deterministic Approach* (pp. 381–394). Youngstown State University. http://eaton.math.rpi.edu/faculty/kramer/mcm/2008mcmsolutions.pdf#page=200

Noti, G. (2017). Do humans play equilibrium?: Modeling human behavior in computational strategic systems. *XRDS: Crossroads, The ACM Magazine for Students*, *24*(1), 29–33. https://doi.org/10.1145/3123740

Pelánek, R. (2014). *Difficulty Rating of Sudoku Puzzles: An Overview and Evaluation*. https://arxiv.org/pdf/1403.7373.pdf

Reeson, C. G., Huang, K., Bayer, K. M., & Choueiry, B. Y. (2007). *An Interactive Constraint-Based Approach to Sudoku*. 1976–1977.

Rosenhouse, J., & Taalman, L. (2012). *Taking Sudoku Seriously: The Math Behind the World's Most Popular Pencil Puzzle*. Oxford University Press.

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Stuart, A. (2007). *The Logic of Sudoku*. Michael Mepham.

Stuart, A. (2008). *SudokuWiki.org*. http://www.sudokuwiki.org/

Stuart, A. C. (2007). *Sudoku Creation and Grading-A Sudoku Creation and Grading*.

Sudoku Dragon: Sudoku Puzzle solving strategies. (n.d.). Retrieved 12 November 2018, from http://sudokudragon.com/sudokustrategy.htm

Sudoku Snake—Solving Techniques. (n.d.). Retrieved 12 November 2018, from http://www.sudokusnake.com/techniques.php

Terveen, L. G. (1995). Overview of human-computer collaboration. *Knowledge-Based Systems*, *8*(2), 67–81. https://doi.org/10.1016/0950-7051(95)98369-H

Van Someren, M., Barnard, Y., & Sandberg, J. (1994). The think aloud method: A practical approach to modelling cognitive. *London: AcademicPress*.