Determining Degree Of Difficulty In Rogo, A TSP-based Paper Puzzle

Dr Nicola Petty, Dr Shane Dye
Department of Management
University of Canterbury
New Zealand
{shane.dye,nicola.petty}@canterbury.ac.nz

Abstract

Rogo®, a pencil and paper puzzle, is based on a subset-selection travelling salesperson problem with a known optimal score. There is an infinite number of possible Rogo puzzles, with at least twelve aspects which may be varied. In order for puzzles to be appealing, they should be difficult enough to be challenging and interesting, but not intractable or tedious. Through examination, mathematical modelling and experimentation on human subjects, we begin research into what elements affect the degree of difficulty of Rogo puzzles. Comparisons are made with other puzzles. Some preliminary results are given.

Key words: Travelling salesperson problem, puzzles.

1 Introduction

Rogos were invented in August 2009 and were developed during 2010, being released as an iPhone app in December 2010. In order to provide a graded level of difficulty which helps people learn and engage with the puzzle, we needed to explore what elements make a Rogo puzzle difficult to solve. This has strong parallels with the problem of instance generation for testing and developing heuristic solution methods.

2 What is Rogo

Rogo is a puzzle based on a prize-collecting, subset selection, Travelling Salesperson Problem set on a rectilinear grid. It was developed in 2009 by Petty and Dye, who have developed an algorithm to solve Rogo puzzles to optimality. Solving a Rogo puzzle involves finding the complete tour of a specified length, avoiding forbidden squares, to maximise the score. In the paper version of the puzzle, the “Best” score is given, along with a “Good” score. These provide targets and stopping criteria, without which the puzzle has limited appeal. All puzzles have a unique solution with regard to the prizes selected, though there may be slightly different loops possible. In Figure 1 a small Rogo is given for illustration, with the solution showing. Rogos can be any size, though the smallest sensible format is 6 by 4 squares, with a loop of length 10, and there seems to be little need to go beyond a total of 150 squares (16 by 9, 12 by 12, 15 by 10) with a loop length of 20.

Rogo has elements in common with Sudoku, mazes and word-search puzzles. Superficially it looks like Sudoku as it is played on a grid of squares, involves numbers, and requires logic to solve. However the similarity ends there. In Sudoku the numbers are merely representative objects, and could as easily be letters or symbols, whereas in

Rogo the values of the numbers are essential to the puzzle. In contrast with Sudoku, the paper version of Rogo requires the player to add, subtract and count. Like a maze, Rogo involves a certain degree of trial and error, and route-finding. Also like a maze, Rogo is completed as soon as the optimal route has been identified and the "Best" score obtained. Rogo is similar to a word-search puzzle in that it requires scanning of the game board, and pattern recognition.

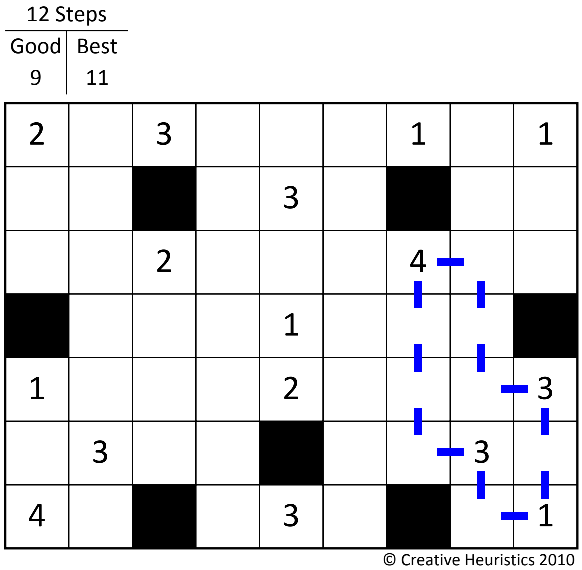


Figure 1: A Rogo puzzle, solved

3 What does a puzzle need to be appealing?

For a puzzle to be appealing and enjoyable, rather than a task, is somewhat personal, but there are some aspects which potentially affect the level of enjoyment. We suggest some of them are degree of difficulty, reward, and aesthetics.

If a puzzle is too easy, it is trivial and solved too quickly and if it is too hard, it can be frustrating and off-putting. This is different for different individuals.

Puzzles need to have an inherent reward. For many puzzles the reward is completion. Satisfaction comes from filling in all the squares in a Sudoku or crossword, or making the completed picture in a jigsaw. Rogo is different from these examples as the goal is not completion; there is no tidy up stage at the end. The sense of accomplishment or discovery comes when the route adds up to the target best score, and you have found the "best" route.

We also suggest that level of difficulty does not map directly onto solution time. Crosswords give a good example of this – some puzzles take a long time because they are large, with many squares. Others may be small and cryptic, and take a long time because of the thinking involved.

It can be suggested that enjoyment is a function of level of difficulty and solution time, and each person will have his or her person "enjoyment function" for a specific type of puzzle. This research focusses on the level of difficulty rather than on the level of enjoyment. Elapsed time to solve is used as a proxy for level of difficulty.

4 Elements of Rogo – what affects level of difficulty?

Twelve different dimensions or aspects of Rogo puzzles have been identified as potentially affecting the degree of difficulty. These are listed below, with a description and explanation as to how each may affect the level of difficulty. It is expected that the different elements would also interact in their effect on solution time.

4.1 Size of grid

The grids vary in size from a total of 24 to 150 squares. It is expected that larger puzzles take longer to solve as there is more area to scan for potential solutions.

4.2 Shape of the grid

The grids may be square or rectangular, landscape orientation or portrait. A narrow puzzle may speed up the search process.

4.3 Placement of black squares

Black (forbidden) squares form obstacles in the Rogo board. Some restrict the number of possible routes. This may make the puzzle instance easier by reducing the number of possible routes, or it can make it more difficult because numbers that look as if they could go together, can't.

4.4 Symmetry

Rotational and reflectional symmetry is often used in Rogo making in order to make the puzzles more pleasing to the eye. The symmetry relates to the placement of the forbidden squares and sometimes the placement of the numerals. This can speed up the search for an optimal solution as route shapes that are identified in one part of the board can be easily replicated in similar parts.

4.5 Number of prizes/density

The more dense the numerals are, the more computation is needed and too many numerals makes the game less interesting, as it is more difficult to get an idea by "eye" of where the high scores are. Blank squares are equivalent to a zero score, but because they are blank it is easier to spot patterns and potential routes in the puzzle.

4.6 Number of steps in the solution

Longer loops require more counting and adding, and have a larger number of possible shapes, thus increasing the solution time.

4.7 Shape of the solution

Shapes that have 'double-back's in them may be harder to find than, for instance, plain rectangles.

4.8 Location of the solution

It could be that people scan from left to right and top to bottom. Thus a solution in the top left would be more quickly found than one in the bottom right.

4.9 Variety in numbers used

If only a small range of numbers is used, this makes the logic and the computation easier. However, the effect of adding one number to a route is fairly consistent, compared with having a range of numbers.

4.10 Magnitude of numbers

Larger numbers make for a greater computational load, and make the marginal effect of adding another single number (reward) to a route greater.

4.11 Number of near misses

As we generate Rogo problem instances, we aim to make sure that there are sufficient routes, distributed throughout the grid, that provide scores close to the optimal. This stops the solution from being trivial.

4.12 Location of largest value rewards

A number noticeably greater in magnitude can draw the eye, and possibly anchor the players thinking. “Surely the 9 must be in the optimal route!” This could affect the solution time either way, depending on whether the high number really is in the optimal route or not.

5 Other work on puzzle difficulty and human performance

Previous research has looked at puzzle difficulty and human performance on puzzle solving. We examined work on Sudoku, TSP as a puzzle and the 15-puzzle.

Research on Sudoku has generally centred on developing a method for classifying puzzle difficulty. One approach for Sudoku is to use algorithms designed to mimic human puzzle solvers. As a logic-based puzzle there are essentially two main processes to use in solving Sudoku puzzles for both humans and computer algorithms. The first is to use logical reasoning to reduce the possible values that can be placed in a cell, isolating those for which only one value is feasible. The second process is trial-and-error, effectively branching over possible cell values. For computer algorithms differences in performance come from how the Sudoku problem is formulated.

For example, Henz and Troung (2009) formulate Sudoku puzzles as satisfiability (SAT) problems and present a tool, called SudokuSat. They classify Sudoku puzzles based on solving times. Chen (2009) formulates Sudoku as a graph colouring problem with edges connecting cells in the same row, column or block. The algorithm proposed uses limited logic levels and branching, where necessary. An index is defined based on the number of feasible values for each cell after limited preprocessing. This index is used to classify puzzle difficulty.

Human performance on the TSP problem has been studied. MacGregor and Ormerod (1996), MacGregor *et al* (2000), and Dry *et al* (2006), among others, tested human performance on visually presented TSP instances. They draw conclusions about the heuristics that might be employed by the problem-solvers and what aspects of the TSP made the problems difficult (in terms of solution time). Two aspects which appeared to impact on difficulty were the number of nodes and the number of nodes on the convex hull. To judge the quality of solutions produced these were compared to solutions found by a (computer) heuristic and to the average length of randomly sampled solutions.

Visually presented TSP instances have a number of features which can complicate conclusions about puzzle difficulty. The subjects generally produced a single solution (without improvement) which was often not the optimal solution. One effect of this is that solution times correspond to different quality solutions. Visual discrimination is also a factor in performance. It was noted by van Rooij *et al* (2003) that studies in which tours were produced by pencil produced proportionally more tours with crossings than those using a computer interface.

Pizlio and Li (2005) studied human performance on the 15-puzzle. They compared the number of solution steps taken with the minimum number required. They formulated a computational model to mimic the apparent human behaviour evident from the results.

6 The Pilot Study

As we are interested in the time that people take to solve Rogo puzzles, we did some preliminary research to see if there were consistencies between people, asking whether there are some puzzles that are more difficult than others for all or nearly all people.

We developed a set of Rogo puzzles which had a range of difficulties, we believed, based on a small sample. Testing sessions were set up, with up to six subjects at a time. The subjects were taught how to solve Rogos and timed as they completed each of the 12 puzzles. The puzzles were checked later to make sure they were correctly completed.

7 Preliminary results

There were 71 subjects in the sample for the pilot study. Summary data is given in Table 1. Scores over 600 seconds were removed. Similarly, puzzles 3, 7 and 12 had a large number of incorrect results or incompletes, so were removed before further analysis.

Puzzle Instance	1	2	3	4	5	6	7	8	9	10	11	12
Solved to Best	50	57	20	51	61	38	7	28	35	22	20	12
Solved to Good	6	6	7	10	5	17	23	14	12	11	12	14
Incorrect	9	6	30	3	1	6	16	11	4	8	3	2
Other	6	2	14	7	3	8	20	10	7	12	8	10
Attempted	71	71	71	71	70	69	66	63	58	53	43	38
Percentage Best	70%	80%	28%	72%	87%	55%	11%	44%	60%	42%	47%	32%
Mean (Best)	140	106	254	173	95	176	327	246	149	176	149	210
Standard deviation	89	82	183	110	89	148	135	173	90	113	107	160

Table 1: Summary data for the 12 puzzle instances.

Puzzle 8 was an interesting in that it had more outliers than the other puzzles. Even some of the “good” puzzlers (ones who solved 8 or more of the 12 puzzles) recorded very long times for puzzle 8. (The puzzle instances are given in the Appendix.)

The nature of the puzzles were examined to see if any could explain some of the variation in time taken to solve. With a sample of only 9 puzzles, this could only give an indication. The number of near misses (routes that scored 1 or 2 less than the ‘best’ score) was a significant predictor, with an R-sq value of 6% and a p-value of 0.000. This would suggest that the higher the number of near misses, the harder the puzzle is, which aligns with aspects of the method used for generating puzzle instances.

The differences in means between the 9 puzzles was examined. The ANOVA resulted in a p value of 0.000. This indicates that there is greater variation in the scores between the different puzzles than there is within the puzzles. Thus we can conclude that there is an element of universality in difficulty, which promises well for further research in this area. Graphs of the times taken for individuals for the puzzles showed little similarity, which indicates that though some puzzles are more difficult than others in general, it is by no means universal to all the puzzle-solvers.

The subjects were classified according to the number of puzzles they were able to solve successfully, with 0 to 4 being "weak", 5 to 7, "medium" and 8 to 11 "good". There was a significant difference in the average time taken for solving between the three groups. Some background information on the individuals, such as their puzzle-solving behaviour and ability at mathematics had no predictive ability on the time taken.

8 What next?

Many of the problems encountered in the paper-based experiment could be reduced or eliminated by the use of the electronic form of Rogo. This would ensure that the puzzles were solved correctly and would require the subject to deliberately choose to give up on a puzzle if they wanted to move on. An electronic form of Rogo would provide a quick and less labour intensive way to collect considerable data on solution times. In addition, the routes created in the course of solving the puzzle could be analysed to explore the kind of thinking used. This, in turn could be possibly be used to inform computer-based heuristic solution methods. The computational aspect of the puzzle is reduced with the electronic medium, as the adding and counting are done by the program. This could be turned off if required to evaluate the impact of computation on solution times.

9 Conclusions

The Rogo puzzle format has a number of aspects that can be controlled to potentially affect degree of difficulty of solving. As a pilot, this study showed that there are many aspects of puzzle-solving related to the nature of the puzzle that can be explored, and there appear to be some general effects, though there are still marked individual differences between people solving the puzzles. This research has the potential to provide interesting insights into both human behaviour, and the nature of puzzles.

References

Chen, Zhe. 2009. "Heuristic reasoning on a graph and game complexity of Sudoku." *Computer Research Repository (CoRR)* arXiv:0903.1659 (http://arxiv.org/abs/0903.1659)

Dry, Matthew, Michael D. Lee, Douglas Vicker and Peter Hughes. 2006. "Human performance on visually presented traveling salesperson problems with varying numbers of nodes." *The Journal of Problem Solving* **1** (1):20–32.

Henz, Martin, and Hoang-Minh Truong. 2009. "SudokuSat—a tool for analyzing difficult Sudoku puzzles." In *Tools and Applications with Artificial Intelligence* – C. Koutsojannis and S. Sirmakessis (eds) SCI **166**: 25–35.

MacGregor, J.N. and T. Ormerod. 1996. "Human performance on the traveling salesman problem." *Perception and Psychophysics* **58** (4):527–539.

MacGregor, J.N., T.C. Ormerod and E.P. Chronicle. 2000. “A model of human performance on the traveling salesperson problem.” *Memory and Cognition* **28** (7): 1183–1190.

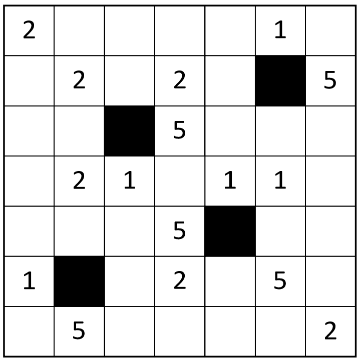
Pizlio, Zygmunt, and Zheng Li. 2005. “Solving combinatorial problems: the 15-puzzle” *Memory and Cognition* **33** (6): 1069–1084.

van Rooij, Iris, Ulrike Stege and Alissa Schactman. 2003. “Convex hull and tour crossings in the Euclidean traveling salesperson problem: implications for human performance studies.” *Memory and Cognition* **31** (2):215–220.

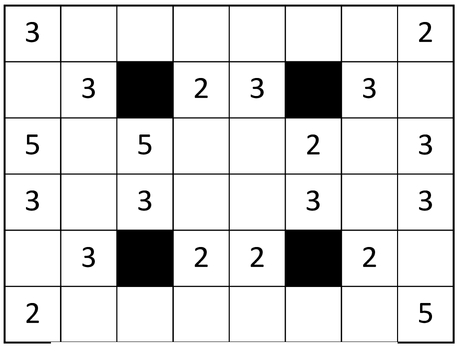
Appendix

The 12 Rogo puzzles used in the experiment.

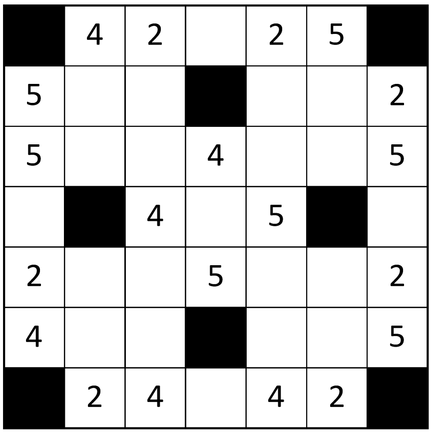
Trial 1 — 12 Steps — Good 19 | Best 21



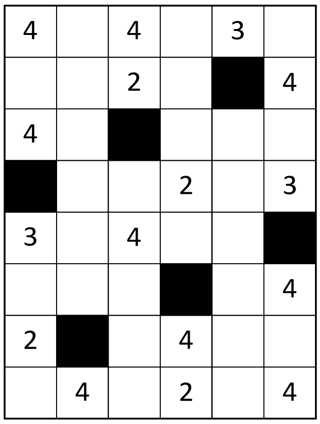
Trial 2 — 12 Steps — Good 22 | Best 25



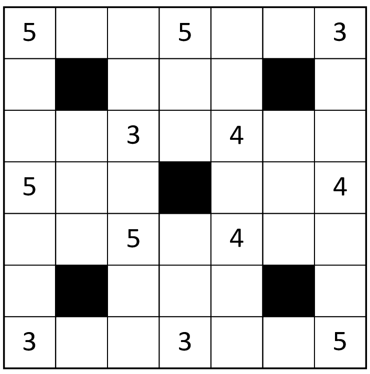
Trial 3 — 12 Steps — Good 22 | Best 23



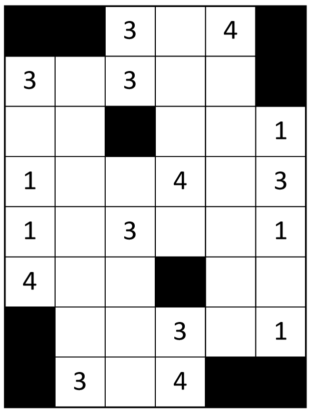
Trial 4 — 12 Steps — Good 16 | Best 19



Trial 5 — 16 Steps — Good 24 | Best 28



Trial 6 — 16 Steps — Good 21 | Best 23



Trial 7 — 16 Steps

Good	Best
28	30

	2	4			2	
5				■		4
	■		5			5
		4	2	2		
2			4		■	
4		■				2
	4			2	4	

Trial 8 — 16 Steps

Good	Best
21	24

■	■		4	4		■	■
■	3					3	■
1							4
	3	■	1	3	■	3	
3			4	3			1

Trial 9 — 16 Steps

Good	Best
25	29

3			3			5
	■			2	■	
	5		3		2	
3	■	2	■	3	■	3
	3		5		2	
	■	2			■	
2			5			5

Trial 10 — 12 Steps

Good	Best
17	18

	4		2	4		4	
4							3
	3	■		2	■		
		■	3		■	3	
2							4
	3		3	4		3	

Trial 11 — 12 Steps

Good	Best
13	15

3			1			3
	3	■		■	2	
		1	3	2		
■	1				3	■
		2	1	1		
	3	■		■	3	
3			3			3

Trial 12 — 16 Steps

Good	Best
22	24

5	1			5	1
2					2
		■	■		
1		5		1	
	2		2		5
		■	■		
2					5
2	2			1	2