Which is harder? - Classification of Happy Cube puzzles

Szilárd András[1], Kinga Sipos[2], Anna Soós[3],
Babeş-Bolyai University, Cluj Napoca, Romania

Motto: *A good puzzle it's a fair thing.*
Ernő Rubik

Abstract. In this paper we study the complexity of the 3D Happy Cube puzzles. The aim is to give the right order of these puzzles. To achieve this goal we studied the cubes from a theoretical point of view (the complexity of solving algorithm), we tested the cubes on a sufficiently large group (in order to obtain an empirical order) and we studied by statistical methods the correlations between the theoretical and the empirical orders. Finally we obtained a practically confirmed theoretical classification of these puzzles which is completely different from the manufacturer's classification.

Keywords: happy cube, complexity, clustering, PCA

Mathematical Subject Classification: 11Y16, 62H25, 62H30, 97A20

Introduction

Happy Cubes are a set of mechanical puzzles created in 1986 by the Belgian toy inventor Dirk Laureyssens and they are also known as "Snafooz" in USA, "Wirrel Warrel" in Netherlands or "Cococrash" in Spain ([1]). The Happy Cube Family is divided into 4 different sets: Little Genius, Happy Cube, Profi Cube, Marble Cube. According to the manufacturer's homepage ([2]) these sets have different difficulties and the difficulty of each set is marked with the corresponding number of stars as follows:

- Little Genius ⋆
- Happy Cube ⋆⋆
- Profi Cube ⋆⋆⋆
- Marble Cube ⋆⋆⋆⋆

Each of these sets contains 6 different models, labeled also with stars accordingly to their difficulty. The Profi Cube and the Marble Cube set consists of the following cubes:

Profi Cube	Confusius	Da Vinci	Marco Polo	Rubens	Watt	Newton
	⋆	⋆⋆	⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆⋆⋆
Marble Cube	Martin L. King	Omar Khayyam	Marie Currie	Buckminster Fuller	Mahatma Gandhi	Albert Einstein
	⋆	⋆⋆	⋆⋆⋆	⋆⋆⋆⋆	⋆⋆⋆⋆⋆	⋆⋆⋆⋆⋆⋆

The Profi Cube set is recommended for the age-group $7-125$ and is considered "Difficult" while the Marble Cube set is recommended for the age-group $9-125$ and is considered "The most difficult."

We have been using these puzzles (the Profi Cube set and the Marble Cube set) for several years not only as an educational toy but also as teaching material in our mathematical modelling activities. These activities were designed for $14-18$ years old students and were focused mainly on developing the concept of graph (and hypergraph), the "rediscovering"

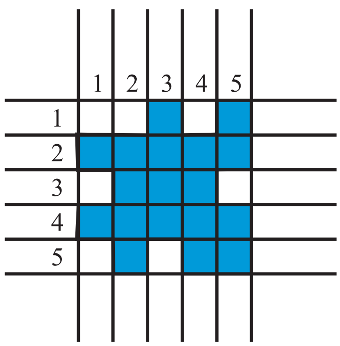
[1]Email address: andraszk@yahoo.com
[2]Email address: kinga_sipos@yahoo.com
[3]Email address: asoos@math.ubbcluj.ro

of the backtracking algorithm and the optimization of backtracking algorithms using cutting conditions. During these activities we (and our students) observed that something is wrong with the difficulty ranking given by the manufacturer. Our impression was that the Marble set is easier than the Profi set and even inside of a set the stars do not contain any relevant information on the difficulty of the model. So the following natural problems arose:

1. How to construct a proper ranking of the cubes by measuring the required time for solving the puzzle (separately for each cube)?
2. How to define a theoretical complexity for these cubes such as the classification determined by this complexity fits the previous ranking?

A little mathematics and algorithm analysis

We represented each face of the cube as an 5×5 matrix with elements 1 and 0. The element in the i^{th} row and j^{th} column is 1 if and only if the corresponding square of the face is full (see figure 1).



	1	2	3	4	5
1	0	0	1	0	1
2	1	1	1	1	1
3	0	1	1	1	0
4	1	1	1	1	1
5	0	1	0	1	1

Figure 1: Representation of a face

To handle the rotations of a face we used 8 such matrices for each face (4 different positions obtained by rotations and the symmetrical of these positions) and we denoted these matrices by $M_1, M_2, \ldots, M_{48}$. Depending on the symmetry of a face among these matrices we can have several identical matrices but this is not a problem. We define a relation $\sim$ on the set of these matrices in the following way: $A \sim B$ if and only if A and B correspond to different pieces and the corresponding pieces can be joined such that the common edge of the pieces is described by the first line of the matrices A and B. Technically the matrices A and B corresponding to different pieces belong to the same equivalence class if and only if

$$
\begin{aligned}
&a_{1i} + b_{1i} = 1, i \in \{2, 3, 4\};\\
&a_{11} + b_{11} \leq 1 \text{ and if } a_{11} + b_{11} = 0, \text{ then } \min\{a_{21}, b_{21}\} = 0;\\
&a_{15} + b_{15} \leq 1 \text{ and if } a_{15} + b_{15} = 0, \text{ then } \min\{a_{25}, b_{25}\} = 0.
\end{aligned}
$$

In this way the solution of a cube can be described by a set of 12 corresponding pairs. For this reason in the first step we generated a 48×48 incidence matrix denoted by F. The element f_{ij} of this matrix is 1 if the corresponding pieces fit each other ($M_i \sim M_j$) and 0 in the other case. Clearly the sum of the j^{th} column in F gives the number of pieces that fit the corresponding piece at a fixed edge. In a solution any piece has its place so, we can start the assembling of the cube from any piece. Hence we can choose the first piece (and position) coded by the matrix M_j such that the number $\sum_{i=1}^{48} f_{ij}$ is minimal. This ensures that even if we do not choose the right pieces at the beginning we have to try a minimal number of possibilities. In a very lucky case if this minimal number is 1, we have a certain starting point. This situation occurs in the case of the Watt Cube (see figure 2 and the other pieces in the appendix).

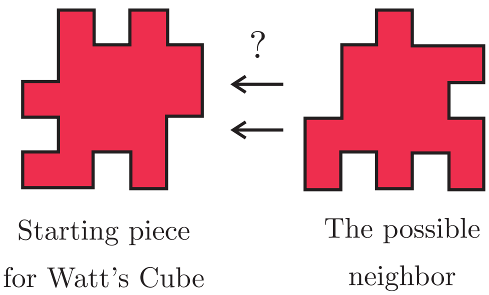


Figure 2: A certain start for the Watt Cube

We denote by c_1 the minimal column sum, so

$$c_1 = \min_{1 \le j \le 48} \sum_{i=1}^{48} f_{ij} \tag{1}$$

is the first complexity indicator. In the following table we have the first complexity indicator for the analyzed cubes:

Profi	Confusius	Da Vinci	Marco Polo	Rubens	Watt	Newton
c_1	8	8	7	6	1	2

Marble	King	Khayyam	Curie	Fuller	Gandhi	Einstein
c_1	2	8	3	2	2	3

Some of the cubes have symmetrical pieces and this must be taken into account. This can be done if in the initial incidence matrix F we introduce fractions. If a piece L (which is coded by the matrices $M_{i_1}, M_{i_2}, \ldots, M_{i_8}$) has some kind of symmetry, than for each M_{i_k} we count the number of matrices in the set $\{M_{i_1}, M_{i_2}, \ldots, M_{i_8}\}$ identical to M_{i_k}. If this number is s_k, than we use $\frac{1}{s_k}$ in the incidence matrix instead of 1 in the rows (and columns) of M_{i_k}. Let's denote by FS the so obtained incidence matrix and by cs_1 the corresponding complexity indicator

$$cs_1 = \min_{1 \le j \le 48} \sum_{i=1}^{48} fs_{ij} \tag{2}$$

For the sake of accuracy we call cs_1 the first corrected complexity indicator.

In the following table we have the first corrected complexity indicator for the analyzed cubes:

Profi	Confusius	Da Vinci	Marco Polo	Rubens	Watt	Newton
cs_1	4	4	4	5	1	2

Marble	King	Khayyam	Curie	Fuller	Gandhi	Einstein
cs_1	2	3	3	2	2	3

In what follows we define recursively 3 more complexity indicators by constructing a graph which illustrates the possibilities we have when assembling a cube. For a better understanding we construct this graph for the Watt Cube. If we start by joining piece 1 and 2, then at one side we have 2 possibilities while on the other side there are 3 different possibilities to continue with a third piece (figure 4). We choose the minimum number of further possibilities and for each such possibility we analyze the configuration we obtain by joining the third piece to the first two. In each step we have a few positions to add the next piece and to each position we can associate the number of pieces which can join the existing construction at that position. If we choose in each step the position with a minimal number of further possibilities, we obtain the graph from figure 5 for the Watt Cube.

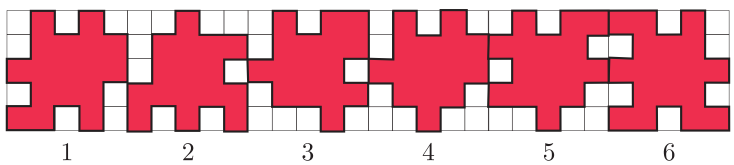


Figure 3: The pieces of the Watt Cube

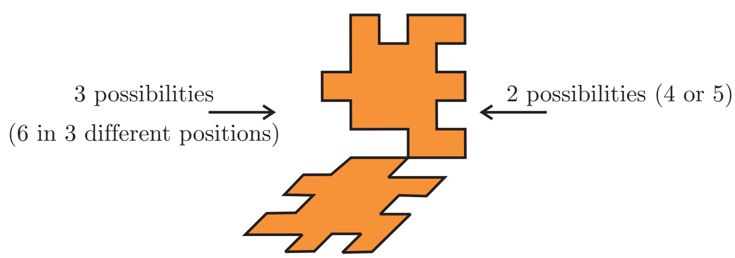


Figure 4: The second step in assembling the Watt Cube

Figure 5 shows that if we proceed by a backtracking algorithm we can assemble the cube in at most 3 attempts (the order of the pieces can be $1-2-5$, $1-2-4-3$ or $1-2-4-5-3-6$).

If we start with an other piece, we obtain more cases (a tree with more leaves), so it is possible (if we are not very lucky) to spend more time assembling the cube. For the choice of the positions we used a kind of Greedy algorithm and it is clear that for some cube this algorithm does not assure that we obtain the simplest tree, but it can be carried out almost effortless (we do not need to memorize configurations). The graph we constructed describes the simplest "Greedy" path to the solution. Appendix 2 contains the corresponding graph for each cube. We measure the hardness of assembling a cube with the complexity of the simplest "Greedy" path. The first corrected complexity indicator is

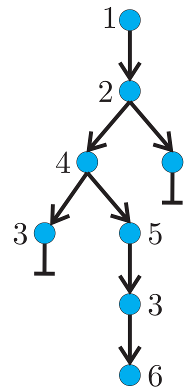


Figure 5: The structure of assembling the Watt Cube

the number of vertices on the first level, hence it is natural to introduce the k^{th} corrected complexity indicator cs_k as the number of vertices on the k^{th} level for $1 \leq k \leq 5$. It is clear that cs_5 is the number of non equivalent solutions, so $cs_5 = 1$ for all the cubes excepting the Omar Khayyam cube for which this indicator is 3. In addition we can count for each cube the number of unsuccessful attempts in the corresponding graph which is the number of leaves that are not on level 5 (in figure 5 this is 2). We denote this number with cs_6 and for each cube we consider the vector $(cs_1^i, cs_2^i, cs_3^i, cs_4^i, cs_5^i, cs_6^i)$, where $1 \leq i \leq 12$ is the number of the cube (in a fixed order). We can construct several complexity numbers from these vectors but in order to obtain a complexity notion connected to the human puzzle solving activity we considered for each $j \in \{1, 2, 3, 4, 5, 6\}$ the order given by the vectors

$$r_j = (cs_j^i)_{1 \leq i \leq 12}.$$

Moreover to obtain a more suitable complexity notion we defined three more complexity numbers as follows:

1. the probability of successful assembling without any backtracking step in the corresponding graph;

2. the average number of steps needed to assemble the cube with backtracking steps in the corresponding graph;

3. the ratio between the number of leaves at the last level and the total number of leaves (this number expresses the conditional probability of the solution relative to the situation when there is no further possibility).

According to these numbers we obtained the following rankings:

	1	2	3	4	5	6	7	8	9	10	11	12
r_7	8	5	6	7	9	12	10	11	1	3	4	2
r_8	6	8	12	5	7	11	9	10	1	3	4	2
r_9	8	5	6	7	12	11	9	10	1	3	4	2

Finally we performed a principal components analysis with the expository variables $(r_j)_{1 \leq j \leq 9}$ and the target vector v obtained from timing the puzzle solving activities of 120 high school student. For calculating the vector v from the initial measurements we used several methods: average times, median times and hierarchical clustering techniques.

Ranking the cubes by puzzle solving activities

We organized several puzzle solving activities for high school students and university students and we measured the time required to assemble each cube. The order of cubes for each participant was generated at random and each participant had to solve 2 cubes before starting the measurements. This introduction ensured that the students got familiar with the cubes, they understood the problem of assembling the cube and they developed their own solving strategy. This can be viewed as a training phase. These two preliminary cubes were not taken from the studied sets. For the training phase we also used a computer program which generated random cube puzzles and had an interactive interface to assemble these. This was necessary to exclude the effect of the learning process on the measurements.

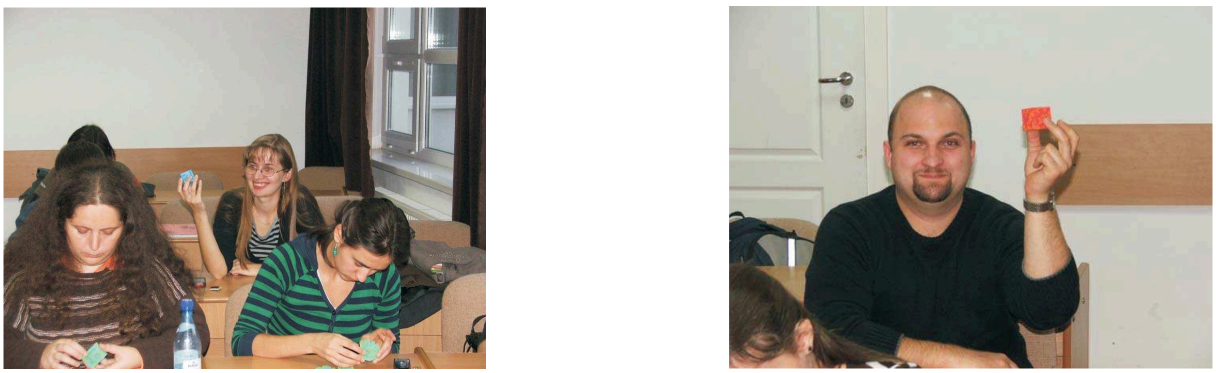


Figure 6: Puzzle solving activities

In this way we obtained a set of measurements m_{ij}, $1 \le i \le 12$, $1 \le j \le n$, where i indicates the number of the cube, j the number of the participant and $n = 120$ was the total number of participants. From these measurements we calculated the following indicators:

1. the average time a_i for each cube;
2. the average time b_i for each cube by neglecting the smallest 10% and the largest 10% of the numbers m_{ij}, $1 \le j \le n$ (for each fixed i);
3. the median time c_i for each cube.

Based on each set of indicators we obtained a possible ranking of the cubes. Finally we used hierarchical clustering to obtain an additional ranking (d_i).

	1	2	3	4	5	6	7	8	9	10	11	12
a_i	8	12	7	11	10	1	6	4	5	3	9	2
b_i	8	12	7	11	6	10	4	1	5	9	3	2
c_i	8	12	7	11	5	6	9	10	1	3	4	2
d_i	7	12	8	5	11	6	1	4	10	9	3	2

The differences between the first two rankings can be explained by the existence of outliers in the measured data. About 30% of the students had a dead point during the

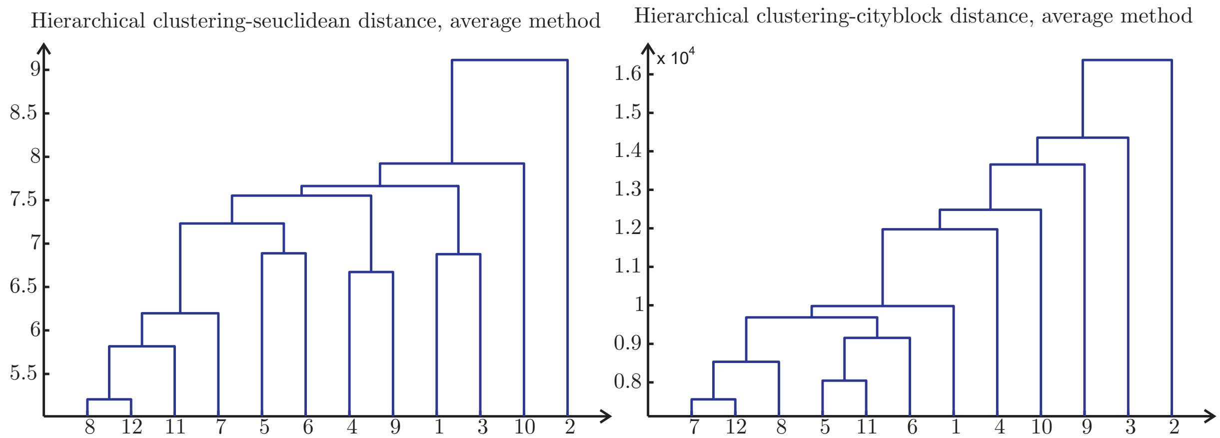


Figure 7: The dendrograms for different distance types and methods

activities and this led to outliers (a cube solved roundly in 1 hour). The fact that the average time for the Watt cube is considerably greater than the average time for the Confusius or for the Rubens cube shows that most of the solvers do not observe the certain start configuration and for this reason they had to try more possibilities. Analyzing the measured times for the Watt cube we can observe that they fit an asymmetric bimodal distribution hence the average is not significant for this cube. These observations and the above table shows that the most natural division into two sets is the following:

I. set (the easy set): 8,12,7,11,5,6

II. set (the hard set): 1,4,10,9,3,2

We denoted by $(a_i)_{1\leq i\leq 12}$, $(b_i)_{1\leq i\leq 12}$, $(c_i)_{1\leq i\leq 12}$ and $(d_i)_{1\leq i\leq 12}$ the above rankings and we performed a factor analysis between the complexity indicators we defined in section 2 and these rankings.

Correlation between theory and practice

By performing a rank correlation analysis (using Kendal and Spearman type coefficients) and a principal components analysis we obtained that:

- the average ranking is best correlated with the ranking obtained from the total number of vertices in the corresponding Greedy graph;
- if we exclude the outliers, the average ranking is best correlated with the ranking obtained from the sum of the first two complexity numbers;
- the median ranking is best correlated with the ranking obtained by using the first corrected complexity number (and the sum of the first two complexity numbers);
- the order obtained from clustering is best correlated with the ranking obtained by using the average number of steps.

In all the cases about 50% of the variance can be explained by a single theoretical ranking and all the theoretical rankings cover $75\% - 85\%$ from the variance of an empirical order.

Concluding remarks

- The rankings obtained from the puzzle solving activities are in accordance with the ranking obtained by evaluating the corresponding graphs constructed by using a Greedy algorithm. Moreover, most of our students confirmed that they used some kind of trial-error, backtracking and Greedy steps in solving these puzzles.

- The categorization used by the manufacturer is completely wrong. The marble set is not the hardest one and the order of cubes used by the manufacturer within the Profi and Marble sets is not correlated with the complexity of the cubes.

- We suggest the following order of the cubes (the cubes of the second set are harder than the cubes of the first set):

  I. set: Omar Khayyam, Mahatma Gandhi, Albert Einstein, Watt, Newton, Martin Luther King

  II. set: Marie Curie, Buckminster Fuller, Confusius, Marco Polo, Da Vinci, Rubens

Acknowledgements

This paper is based on the work within the project Primas. Coordination: University of Education, Freiburg. Partners: University of Genève, Freudenthal Institute, University of Nottingham, University of Jaen, Konstantin the Philosopher University in Nitra, University of Szedged, Cyprus University of Technology, University of Malta, Roskilde University, University of Manchester, Babeş-Bolyai University, Sør-Trøndelag University Colleage.

Our activities were part of the FP7 project PRIMAS[4] while the modelling activities leading to the problem of classification were part of the Comenius project DQME II[5]. We are very grateful to our colleagues Örs Nagy (Babeş-Bolyai University, Faculty of Psychology and Education), Hajnalka Csapó (Márton Áron Highschool, Miercurea Ciuc), Jutka Szilágyi (Báthory István Highschool, Cluj Napoca), Emőke Szilágyi, István Mátéfi (Bolyai Farkas Highschool, Târgu Mureş) and Andrea Bartos-Kocsis for their help in designing the puzzle solving activities and for their constructive observations. The first author's research was partially supported by the Hungarian University Federation of Cluj Napoca.

Appendix 1

In figures 8 and 9 we illustrated all the possible (non equivalent) solutions of the Profi and Marble cubes. The Omar Khayyam cube is the only cube with several non equivalent solutions.

Appendix 2

Figures 10 and 11 illustrate the constructed graph for each cube.

[4]Promoting inquiry in Mathematics and science education across Europe, Grant Agreement No. 244380

[5]Developing Quality in Mathematics Education, for more details see http://www.dqme2.eu/

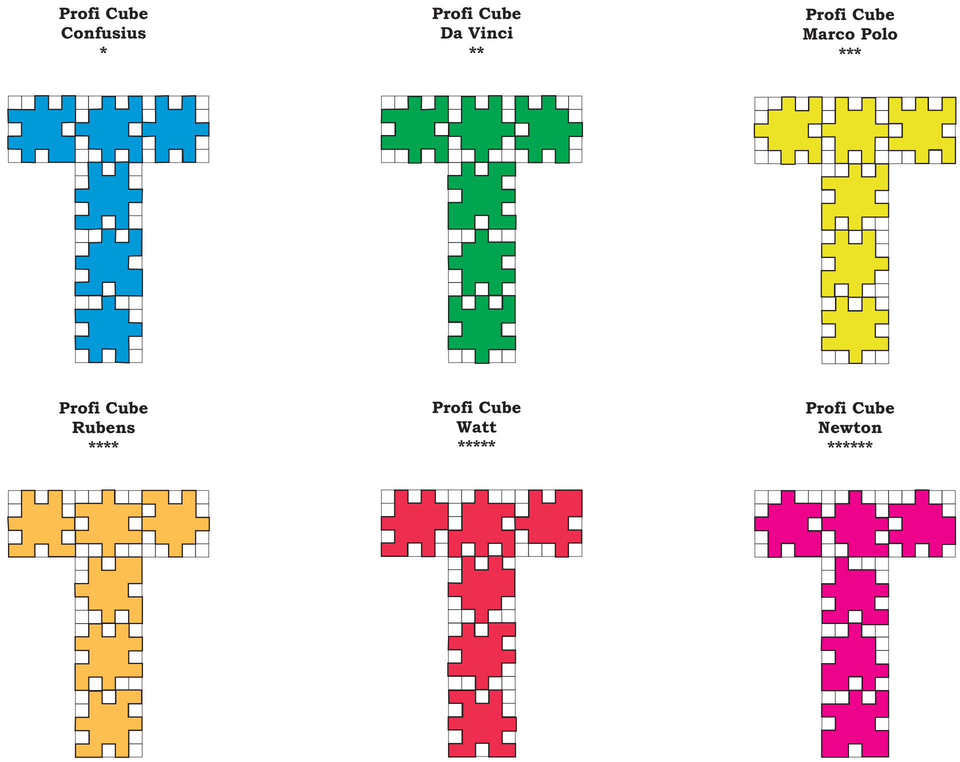


Figure 8: Solutions for the Profi Cube

References

[1] http://en.wikipedia.org/wiki/Happy_Cube

[2] The official site of Happy Cube: http://www.happycube.com/

[3] Brian S. Everitt: Cluster Analysis, Wiley, 4th edition (January 20, 2009)

[4] I.T. Jolliffe: Principal Component Analysis, Springer, 2nd edition (October 1, 2002)

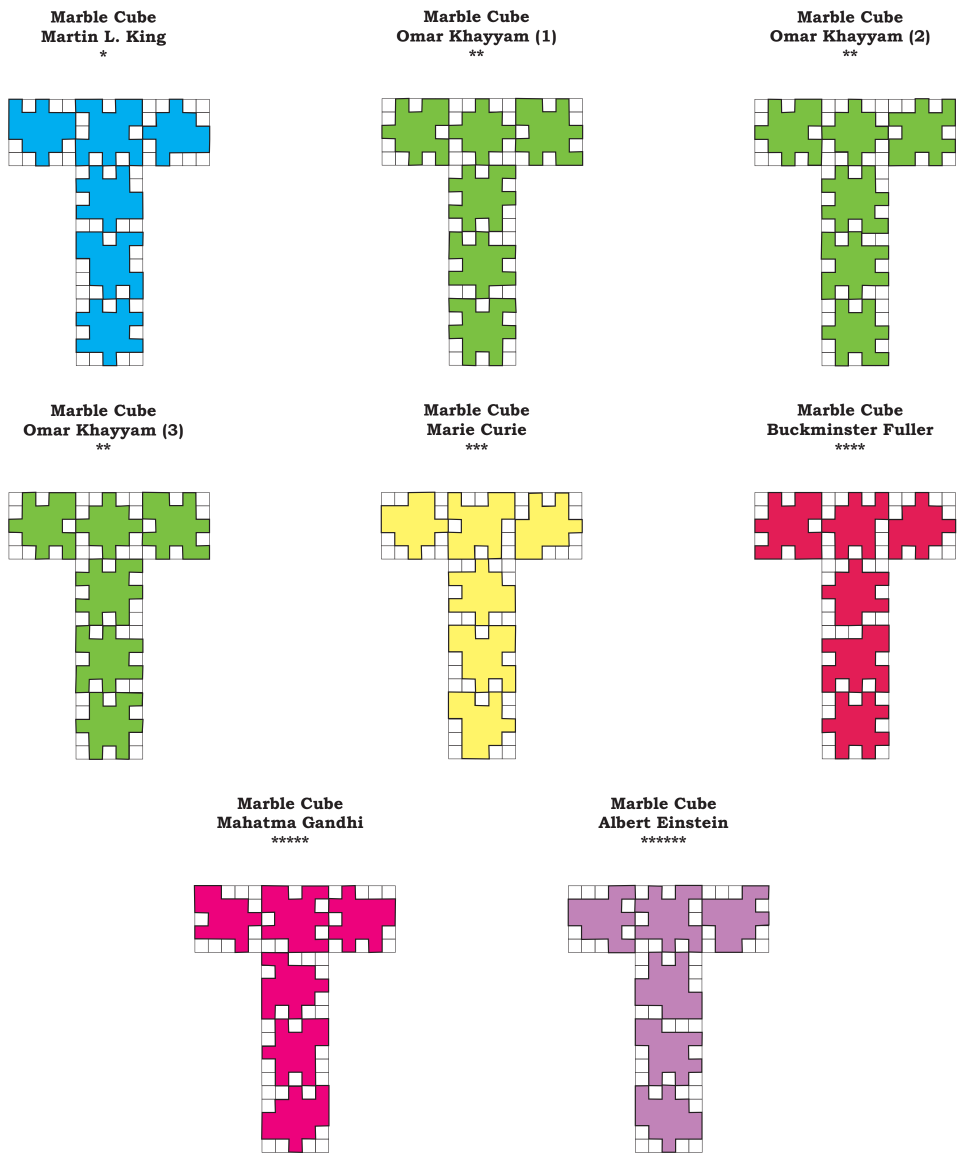


Figure 9: Solutions for the Marble Cube

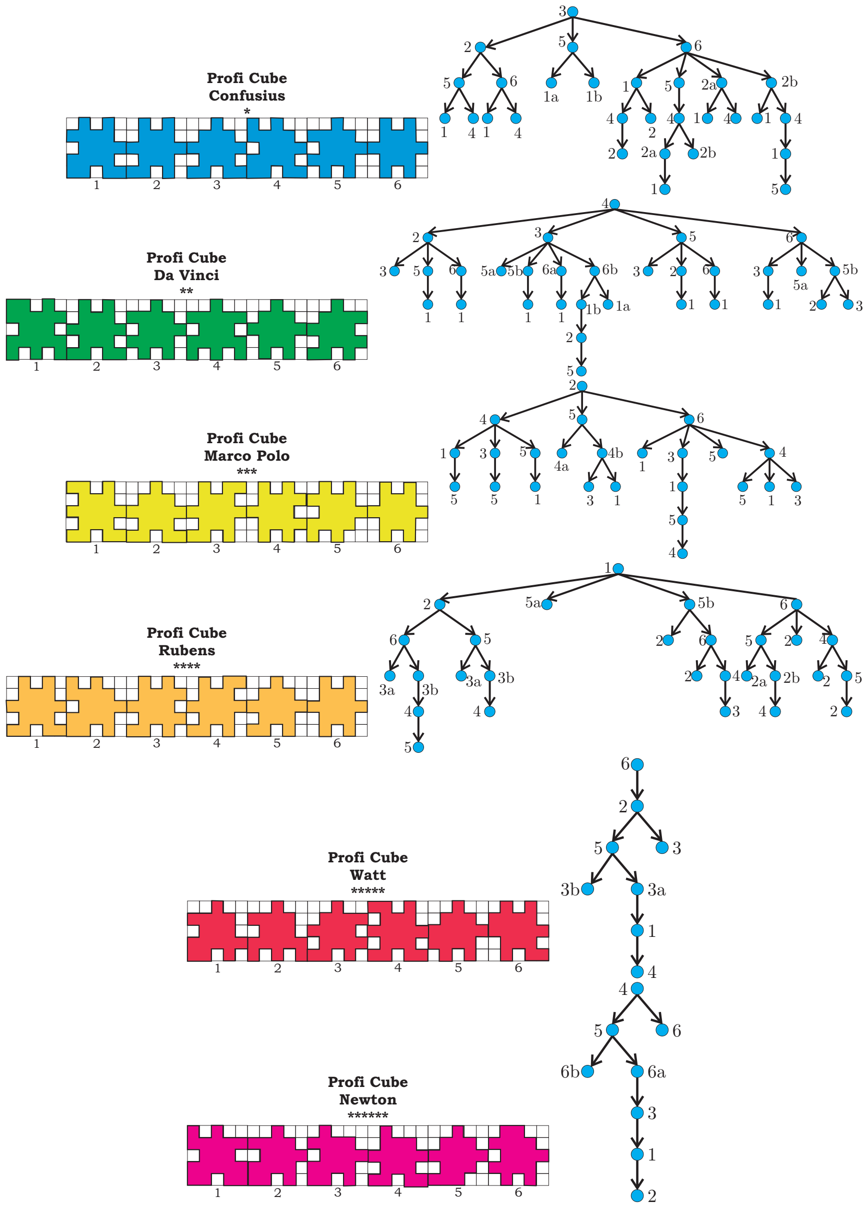


Figure 10: The path to the solution for the Profi set

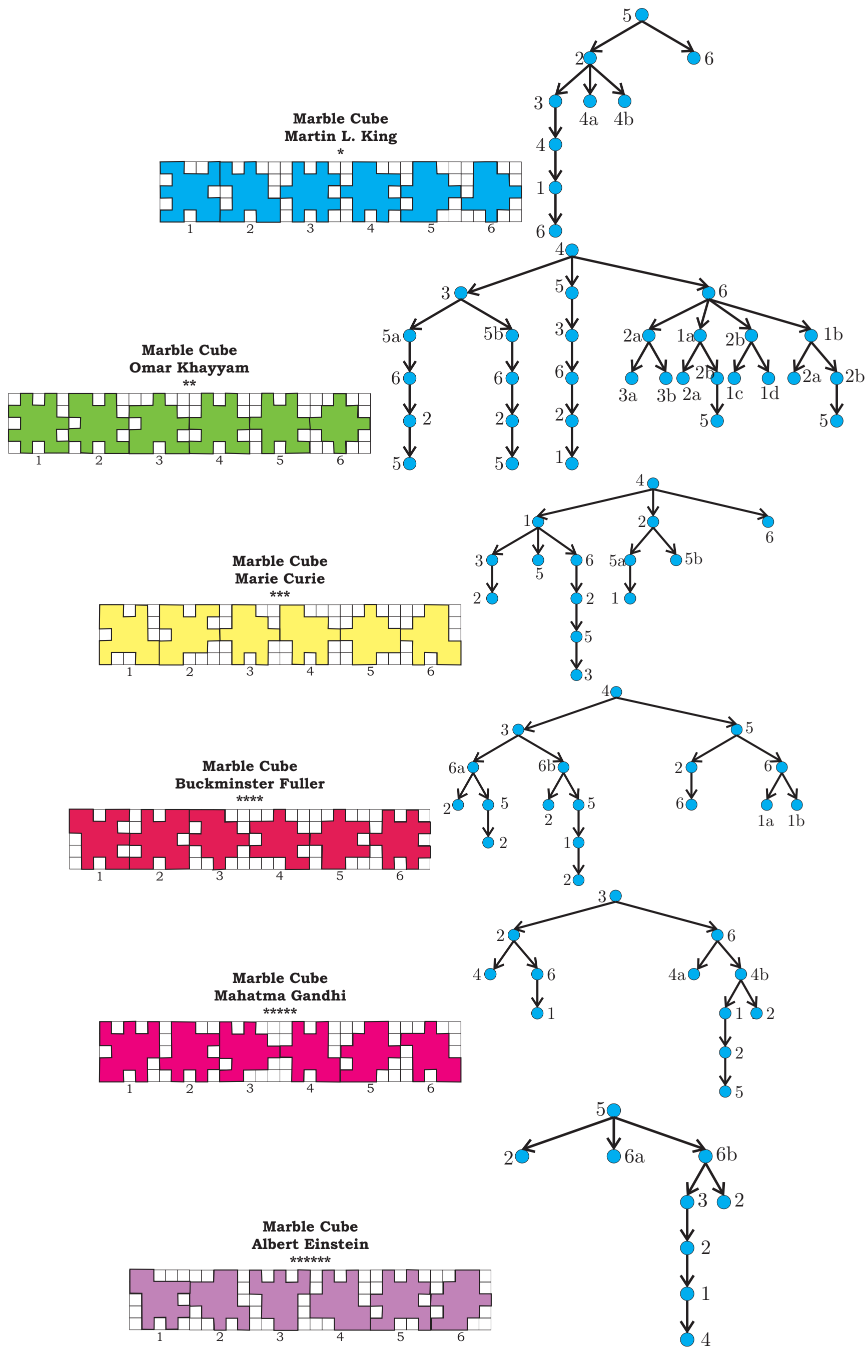


Figure 11: The path to the solution for the Marble set