A Demigod's Number for the Rubik's Cube

Arturo Merino[†] Bernardo Subercaseaux[‡]

[†]Universidad de O'Higgins, Rancagua, Chile
[‡]Carnegie Mellon University, Pittsburgh, USA

Abstract

It is by now well-known that any state of the $3 \times 3 \times 3$ Rubik's Cube can be solved in at most 20 moves, a result often referred to as *"God's Number"*. However, this result took Rokicki et al. around 35 CPU years to prove and is therefore very challenging to reproduce. We provide a worse bound of 36 moves, but that has two main advantages: (i) it is easy to reproduce, verify, and can be presented in one or two lectures, and (ii) our main idea generalizes to bounding the diameter of other vertex-transitive graphs by at most twice its true value, hence the name *"demi-god number"*. Our approach, however, does not actually *prove the bound*, but rather *proves that the bound is overwhelmingly likely to be true*, in a similar sense to how a probabilistic primality test can prove that a number is overwhelmingly likely to be prime. We thus believe that these ideas can lead to interesting philosophical discussions about *plausibility* in the classroom.

We turn the cube and it twists us.

Ernő Rubik.

1 Introduction

The $3 \times 3 \times 3$ Rubik's Cube, illustrated in Figure 1, is arguably one of the most iconic puzzles ever created, and one of the best-selling toys of all time; its beautiful balance of simplicity (it only has 6 faces, with 54 colored stickers) and complexity (it has over 4.3×10^{19} possible states) has captured the attention of millions of people around the world since its invention in 1974. Naturally, such a combinatorially rich puzzle has raised a variety of interesting mathematical questions, with the most famous one being:

What is the minimum number of moves required to solve the Rubik's Cube from any starting position?

To make this question precise, we consider the *half-turn metric*, in which turning either of the 6 faces of the cube by any amount (i.e., 90°, 180°, or 270°) counts as a single move.

After a series of incremental improvements detailed in Table 1, Rokicki et al. [23] proved that 20 moves are always enough to solve any Rubik's cube in the half-turn metric, a result often referred to as *"God's Number"*. This result, however, was obtained by a mixture of mathematical ideas and extensive computation, taking around 35 CPU years. As a result, verifying the correctness of the so-called God's Number is extremely challenging, and the required computations have likely never been reproduced independently. While the result is widely believed to be true, our goal is to provide an alternative approach (providing a weaker bound) that can be easily understood and reproduced in e.g., a classroom setting.

Table 1: Historical bounds on the maximum number of moves required to solve the Rubik's Cube [21].

Year	Lower bound	Upper bound
1981	18	52
1990	18	42
1992	18	39
1992	18	37
1995	18	29
1995	20	29
2005	20	28
2006	20	27
2007	20	26
2008	20	25
2008	20	23
2008	20	22
2010	20	20

1.1 Summary

Our approach is based on the following observation: in any vertex-transitive graph (i.e., a graph where every vertex *"looks the same"*), the diameter D (i.e., the maximum distance between any two vertices) is at most twice the mean distance between vertices, m. This observation corresponds to a question posed by Alan Kaplan [25], which was answered in the more general context of homogeneous compact metrics spaces by Herman and Pakianathan [11]. Our proof, however, is elementary and self-contained, and we believe it can be a nice addition to a first course on graph theory.

Applying the previous observation to the Cayley graph of the Rubik's cube (defined formally in Section 2, this graph has the possible states of the cube as vertices and edges between pair of states that are one "move" away from each other), if we knew the average distance μ between states of the Rubik's Cube, we could bound the diameter of the Rubik's Cube by at most twice this value. While we cannot directly compute the exact

value of μ for the Rubik's Cube[1], we can provide a good estimate by sampling random pairs of states and computing an upper bound on their distance through standard Rubik's algorithms (i.e., Kociemba's *Two-Phase Algorithm* [17]). Each of these upper bounds for the distance between a pair of states is certified by a short sequence of moves, so one can trust the result without needing to trust the algorithm. Through simple concentration bounds, we will argue that the empirical average of $\widehat{\mu} \approx 18.3189$ we obtain is *overwhelmingly likely* to be a good estimate of the true mean distance μ, and therefore an audience should reasonably trust that the diameter of the Rubik's Cube is at most 36 (since it must be an integer). As a first step, we will prove the following theorem.

Theorem 1. *Given a state s of the Rubik's Cube, let $d(s)$ be the distance from s to the solved state. Let S be a set of states of the Rubik's cube sampled uniformly at random, and let $\widehat{\mu}_S = \frac{1}{|S|}\sum_{s\in S} d(s)$ be the random variable corresponding to the average distance between states in S and the solved state. Then, if D denotes the diameter of the Rubik's Cube, we have*

$$\Pr_S\left[D \geq 2\widehat{\mu}_S + 0.36\right] < 2\exp\left(-\frac{|S|}{1\,541\,939}\right).$$

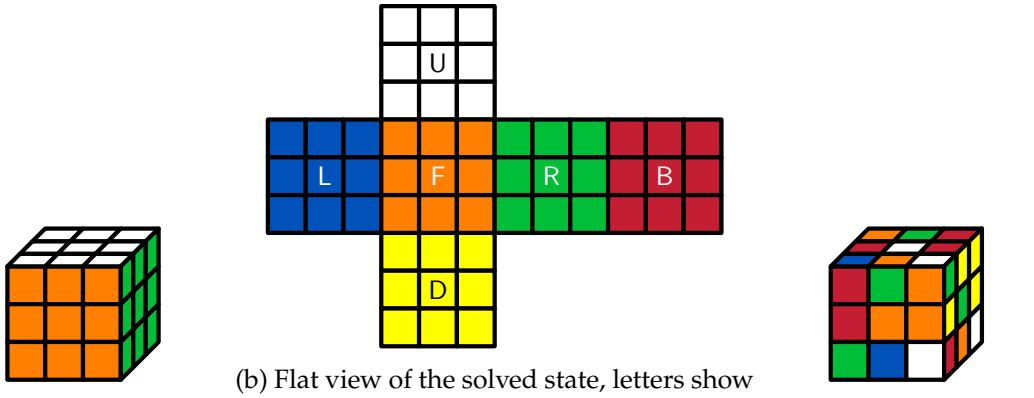
Note immediately that Theorem 1 implies that if $D > 36$ (and thus $D \geq 37$ since the diameter is an integer), the probability of obtaining $\widehat{\mu}_S \approx 18.318$ for $|S| = 10^7$ is less than 10^{-7}. However, this experience can be observed repeatedly, which is therefore a great degree of probabilistic evidence for $D \leq 36$. Unfortunately, Theorem 1 is somewhat computationally expensive, as it requires the number of samples $|S|$ to be around 10 million if we desire a probability of error under 10^{-7}. As we show in Section 4, this takes roughly 100 hours of computation. In order to see how to reduce the required computation, let us briefly discuss how Theorem 1 is obtained. To obtain an upper bound on D, we can leverage the $D < 2\mu$ observation and look for an upper bound on μ, which we can do with a probabilistic guarantee by considering an empirical average μ_S. However, a priori it could be that the true average μ is much larger than our empirical estimate μ_S due to a small number of states that are very far and we are not likely to sample randomly; a "long tail" phenomenon. Our solution to this problem is using a *"Human's number"*, which is an unconditional modest upper bound on D. For instance, the so-called *"beginner's method"* suffices to obtain an upper bound of 205 moves.

Lemma 1 (Human's Number)**.** *Any position of the Rubik's cube can be solved in at most 205 moves.*[2]

Thus, in general, our work can be interpreted as a method for transforming a *"Human's number"* into a *"Demi-god number"* that is easy to trust and at most twice the real *"God's number"*. Now, going back to the problem of how to reduce the number of samples, the 1 541 939 constant in Theorem 1 is a consequence of the constant 205 in Lemma 8. We can improve on this by showing first that in a large percetange of the cases, we can use an

[1]Recall that it has over $4 \cdot 10^{19}$ many states.

[2]A proof sketch is provided in the appendix. In general, we expect anyone familiar with the beginner's method to find this bound trivial, since it is far from being tight.



(a) Solved state

(b) Flat view of the solved state, letters show how we will refer to faces (see Appendix A)

(c) A scrambled state

Figure 1: Illustration of the $3 \times 3 \times 3$ Rubik's Cube.

upper bound of 20 instead of 205. Indeed, let us say that a state is *"far from being solved"* if it requires strictly more than 20 moves to be solved. Naturally, the "God's number" result corresponds to the inexistence of any state that is "far from being solved", but proving this requires a great computational effort. Instead, we use a much simpler argument: if the proportion of states that are "far from being solved" were to be bigger than, say, 0.03%, then we would certainly expect to see a state that is "far from being solved" after 50 000 random samples. Yet, experimentally we do not see any such state after 500 000 samples, which makes the possibility of more than 0.03% of total states being "far from being solved" extremely slim. This way, we can separate μ into:

$$\left[p_{\text{far}} \cdot \sum_{s \text{ far from being solved}} d(s)\right] + \left[(1 - p_{\text{far}}) \cdot \sum_{s \text{ not far from being solved}} d(s)\right]$$

Finally, we believe that our approach can be used as an example to motivate philosophical conversations about *plausibility*, a notion explored at large by George Polya in his book *"Mathematics and Plausible Reasoning"* [20]. In a nutshell, the idea is that while certain types of reasoning do not provide full proof of a statement, they can provide a high degree of confidence in its truth [4]. This is the case, for example, with probabilistic primality tests, which may allow us to confidently assert that a number with hundreds of millions of digits is prime, without providing a proof in the traditional sense of the word.

2 Preliminaries

Let us introduce the notation and definitions required to work over the Rubik's cube mathematically. We will see the Rubik's cube as a group, a standard idea (see e.g., [6, 3]) that we make explicit nonetheless to make our work as self-contained as possible.

Let S_n denote the group of permutations of n elements under the composition opera-

tion, denoted by $\circ$. For example, $(2,3,1) \in S_3$ is the permutation defined by:

$$1 \mapsto 2,\ 2 \mapsto 3,\ 3 \mapsto 1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix},$$

and

$$(2,3,1) \circ (1,3,2) = (2,1,3).$$

We identify each state of the Rubik's cube with a permutation of its stickers, of which there are $6 \cdot 9 = 54$ (9 per each of the 6 faces), and as moving faces in the Rubik's cube results in a permutation of its stickers, the Rubik's cube can be seen as a subgroup of S_{54}, as illustrated in Figure 2.[3] Furthermore, as no moves affect the relative position of the center stickers $(5, 14, 23, 32, 41, 50)$, we can see the Rubik's cube as a subgroup of S_{48}. To complete the definition of the Rubik's cube group, we need to define the "moves", which correspond to the following permutations (omitting the values that are not affected by the move):

$$\mathbf{R} = \begin{pmatrix} 3 & 6 & 9 & 21 & 24 & 27 & 28 & 29 & 30 & 31 & 33 & 34 & 35 & 36 & 48 & 51 & 54 \\ 48 & 51 & 54 & 3 & 6 & 9 & 30 & 33 & 36 & 29 & 35 & 28 & 31 & 34 & 21 & 24 & 27 \end{pmatrix},$$

$$\mathbf{L} = \begin{pmatrix} 1 & 4 & 7 & 10 & 11 & 12 & 13 & 15 & 16 & 17 & 18 & 19 & 22 & 25 & 46 & 49 & 52 \\ 19 & 22 & 25 & 12 & 15 & 18 & 11 & 17 & 10 & 13 & 16 & 46 & 49 & 52 & 1 & 4 & 7 \end{pmatrix},$$

$$\mathbf{U} = \begin{pmatrix} 1 & 2 & 3 & 4 & 6 & 7 & 8 & 9 & 19 & 20 & 21 & 28 & 29 & 30 & 37 & 38 & 39 \\ 3 & 6 & 9 & 2 & 8 & 1 & 4 & 7 & 10 & 11 & 12 & 19 & 20 & 21 & 28 & 29 & 30 \end{pmatrix},$$

and similarly, **D**, **F**, and **B** can be deduced from Figure 2. The moves **R'**, **L'**, **U'**, etc., correspond to the inverse of the moves **R**, **L**, **U**, etc., respectively. A sequence of moves is simply a composition of moves, omitting the composition symbol, e.g., **R U**2 corresponds to the permutation $\mathbf{R} \circ \mathbf{U} \circ \mathbf{U}$.

Let $\mathcal{R}$, the Rubik's cube group, be the subgroup of S_{54} generated by the moves, i.e.,

$$\mathcal{R} = \langle \mathbf{R}, \mathbf{L}, \mathbf{U}, \mathbf{D}, \mathbf{F}, \mathbf{B} \rangle,$$

which is well-defined since $\mathbf{M}^3 = \mathbf{M'} = \mathbf{M}^{-1}$ for every move $\mathbf{M} \in \{\mathbf{R}, \mathbf{L}, \mathbf{U}, \mathbf{D}, \mathbf{F}, \mathbf{B}\}$. As described with words above, $\mathcal{R}$ is clearly isomorphic to a subgroup of S_{48} due to the centers of each face being unaffected by the generators. Finally, note that this group perspective blurs the difference between sequences of moves and states of the cube; every move corresponds to a state of the cube (the result of applying the move to the solved state), and every state reachable from the solved state by moves corresponds to an equivalence class of all move sequences reaching that state.

Definition 1 (Cayley Graph)**.** *Given a group $G = (A, \star)$ and a set of generators $S \subseteq A$, the* Cayley graph *G_S is the graph whose vertex set is A and where two vertices $u, v \in G$ are adjacent if there exists a generator $s \in S$ such that $u \star s = v$.*

[3] More in general, a classic theorem of Cayley states that any group is isomorphic to a subgroup of a symmetric group [14, Section 1.3].

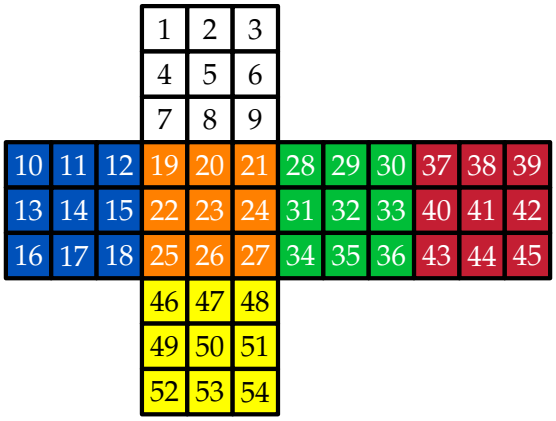


Figure 2: Illustration of the Rubik's cube as a subgroup of S_{54} (or even S_{48} due to the centers being static).

We will use $G_{\mathcal{R}}$ to denote the Cayley graph of the Rubik's cube group under the following set of generators:

$$S_{\mathcal{R}} := \{\mathsf{R}, \mathsf{R'}, \mathsf{R2}, \mathsf{L}, \mathsf{L'}, \mathsf{L2}, \mathsf{U}, \mathsf{U'}, \mathsf{U2}, \mathsf{D}, \mathsf{D'}, \mathsf{D2}, \mathsf{F}, \mathsf{F'}, \mathsf{F2}, \mathsf{B}, \mathsf{B'}, \mathsf{B2}\}.$$

Note that the reason we include e.g., **R'** and **R2** in $S_{\mathcal{R}}$ is that we want to count those as single moves of the cube. Counting **R2** as 2 moves leads to a different metric, usually called the *Quarter-turn Metric*, where *God's number* is 26 [22].

Definition 2 (Diameter)**.** *The* diameter D *of a graph* $G = (V, E)$ *is the maximum distance between any two vertices, where the distance between two vertices* $u, v \in V$ *is the length of the shortest path between them. That is,*

$$D = \max_{u,v \in \binom{V}{2}} d(u, v).$$

In this language, "God's number" is simply the diameter of the Rubik's cube graph $G_{\mathcal{R}}$.

Definition 3 (Mean distance)**.** *The* mean distance μ *of a graph* $G = (V, E)$ *is the average distance between any two vertices, that is,*

$$\mu = \frac{1}{\binom{|V|}{2}} \sum_{u,v \in \binom{V}{2}} d(u, v).$$

Definition 4 (Graph automorphism)**.** *An* automorphism *over a graph* $G = (V, E)$ *is a bijection* $\varphi : V \to V$ *such that for any two vertices* $u, v \in V$*, we have that* $(u, v) \in E$ *if and only if* $(\varphi(u), \varphi(v)) \in E$*.*

Using the previous definition repeatedly leads to the following trivial lemma.

Lemma 2. *If φ is an automorphism over a graph $G = (V, E)$, then for any two vertices $u, v \in V$, we have that $d(u, v) = d(\varphi(u), \varphi(v))$.*

Definition 5 (Vertex Transitivity)**.** *A graph $G = (V, E)$ is* vertex-transitive *if for any two vertices $u, v \in V$, there exists an automorphism φ such that $\varphi(u) = v$.*

We now state a folklore idea that will be key to our analysis of the Rubik's cube graph.

Lemma 3. *Every Cayley graph is vertex-transitive, and in particular, the Rubik's cube graph $G_{\mathcal{R}}$ is vertex-transitive.*

Proof. Let $G = (A, \star)$ be a group and $S \subseteq A$ a set of generators. We must prove that there is an automorphism φ such that $\varphi(u) = v$ for any two vertices $u, v \in A$. Let us define

$$\varphi : A \to A, \quad \varphi(x) = v \star u^{-1} \star x.$$

This definition directly implies $\varphi(u) = v$, and φ is clearly bijective since $x \mapsto u \star v^{-1} \star x$ is an inverse for φ. It remains to prove that for any two vertices $x, y \in A$, we have that $(x, y) \in E$ if and only if $(\varphi(x), \varphi(y)) \in E$. Where, by definition of the Cayley graph, a pair of vertices (a, b) is in E if there exists a generator $s \in S$ such that $a \star s = b$, and equivalently, if $a^{-1} \star b \in S$. Now observe that

$$\begin{aligned}
\varphi(x)^{-1} \star \varphi(y) &= \left(v \star u^{-1} \star x\right)^{-1} \star (v \star u^{-1} \star y) \\
&= (x^{-1} \star u \star v^{-1}) \star (v \star u^{-1} \star y) \\
&= (x^{-1} \star u) \star (v^{-1} \star v) \star (u^{-1} \star y) \\
&= x^{-1} \star (u^{-1} \star u) \star y = x^{-1} \star y,
\end{aligned}$$

from where we conclude by noting that

$$\begin{aligned}
(x, y) \in E &\iff x^{-1} \star y \in S \\
&\iff \varphi(x)^{-1} \star \varphi(y) \in S \\
&\iff (\varphi(x), \varphi(y)) \in E.
\end{aligned}$$

□

We conclude this section with a simple lemma stating that in a vertex-transitive graph, given that all nodes are "essentially the same", we can think of the mean distance as the average distance from a fixed node, instead of between all pairs.

Lemma 4. *Let x be any vertex in a vertex-transitive graph G. Then we have $\mu = \frac{\sum_{v \in V} d(x,v)}{|V|-1}$.*

Proof. First, note that by definition of mean distance we have

$$\mu = \frac{1}{\binom{|V|}{2}} \sum_{u,v \in \binom{V}{2}} d(u, v) = \frac{1}{|V|(|V|-1)} \sum_{u \in V} \sum_{v \neq u \in V} d(u, v).$$

Because of vertex-transitivity, for any vertex $u \in V$, there exists an automorphism φ_u such that $\varphi_u(u) = x$. Therefore, using Lemma 2 we have

$$\mu = \frac{1}{|V|(|V|-1)} \sum_{u \in V} \sum_{v \neq u \in V} d(u,v) = \frac{1}{|V|(|V|-1)} \sum_{u \in V} \sum_{v \neq u \in V} d(x, \varphi_u(v))$$

But as $\varphi_u : V \to V$ is a bijection for every u, we have

$$\sum_{v \neq u \in V} d(x, \varphi_u(v)) = \sum_{v \neq x \in V} d(x,v) = \sum_{v} d(x,v),$$

and thus

$$\mu = \frac{1}{|V|(|V|-1)} \sum_{u \in V} \sum_{v \in V} d(x,v) = \frac{|V|}{|V|(|V|-1)} \sum_{v \in V} d(x,v) = \frac{\sum_{v \in V} d(x,v)}{|V|-1}. \qquad \square$$

3 The Relationship Between Diameter and Mean Distance

In this section, we explore the relationship between the diameter and the mean distance of a graph, and show that for vertex-transitive graphs the diameter is at most twice the mean distance. While this result is implied by [11], we offer a more elementary exposition.

First, let us note that in arbitrary graphs, the diameter D can be much larger than the mean distance μ.

Proposition 1 (Folklore, cf. [26])**.** *For every n, there are graphs on n vertices such that $D/\mu = \Omega(n^{1/2})$.*

Proof. We can construct a graph G by taking a clique on n vertices and attaching to it a path on $n^{1/2}$ vertices, as illustrated in Figure 3.

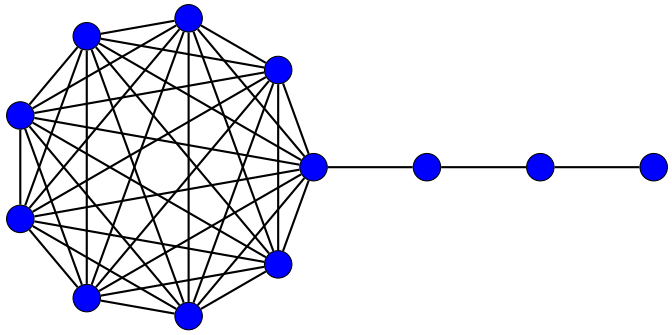


Figure 3: A graph G with $n = 9$, illustrating the proof of Proposition 1.

The diameter of this graph is $D = n^{1/2} + 1$. Noting that $|V| = n + n^{1/2} = \Omega(n)$, the

mean distance can be calculated as follows:

$$\begin{aligned}\mu &= \frac{1}{\binom{|V(G)|}{2}} \sum_{u,v \in \binom{V(G)}{2}} d(u,v) \\ &= \frac{1}{\Omega(n^2)} \Bigg(\underbrace{1 \cdot O(n^2)}_{\text{between clique vertices}} + \underbrace{O(n^{1/2}) \cdot O(n)}_{\text{between path vertices}} + \underbrace{O(n^{1/2}) \cdot n \cdot n^{1/2}}_{\text{clique-to-path}} \Bigg) \\ &= O(1). \end{aligned}$$

□

Furthermore, Wu et al. proved that this bound is asymptotically tight, meaning that $D/\mu = O(n^{1/2})$ [26]. It turns out, however, that such a gap between D and μ is not possible in vertex-transitive graphs, where D and μ are always a constant factor away. Before presenting a short proof, let us present the main intuition, which is illustrated in Figure 4. The main idea is that in a vertex-transitive graph, if we pick two diametrically-opposed vertices u, v, i.e., $d(u,v) = D$, and then consider increasingly large "balls" centered at u and v (i.e., a ball $B(r,x)$ of radius r centered at a vertex x is the set of vertices at distance at most r from x), then at some point r those balls will intersect for the first time. Therefore, there exists some vertex $w \in B(u,r) \cap B(v,r)$ that is at most r away from u and at most r away from v. We thus obtain

$$D = d(u,v) \leq d(u,w) + d(w,v) \leq 2r,$$

using the triangle inequality. Now, note that more than half the vertices must be outside $B(u, r-1)$, since $|B(v, r-1)| = |B(u, r-1)|$ by vertex transitivity, and $|B(v, r-1)| + |B(u, r-1)| \leq n$, since by definition r is the smallest at which those balls intersect. We are therefore saying that around half the vertices are at distance greater or equal than r from u, and as $r \geq D/2$, that makes the average distance from u (and thus the mean distance μ by Lemma 4) at the very least $D/4$. As our proof shows next, the right bound is actually $D/2$. A similar observation is used by [19, Proposition 3.4].

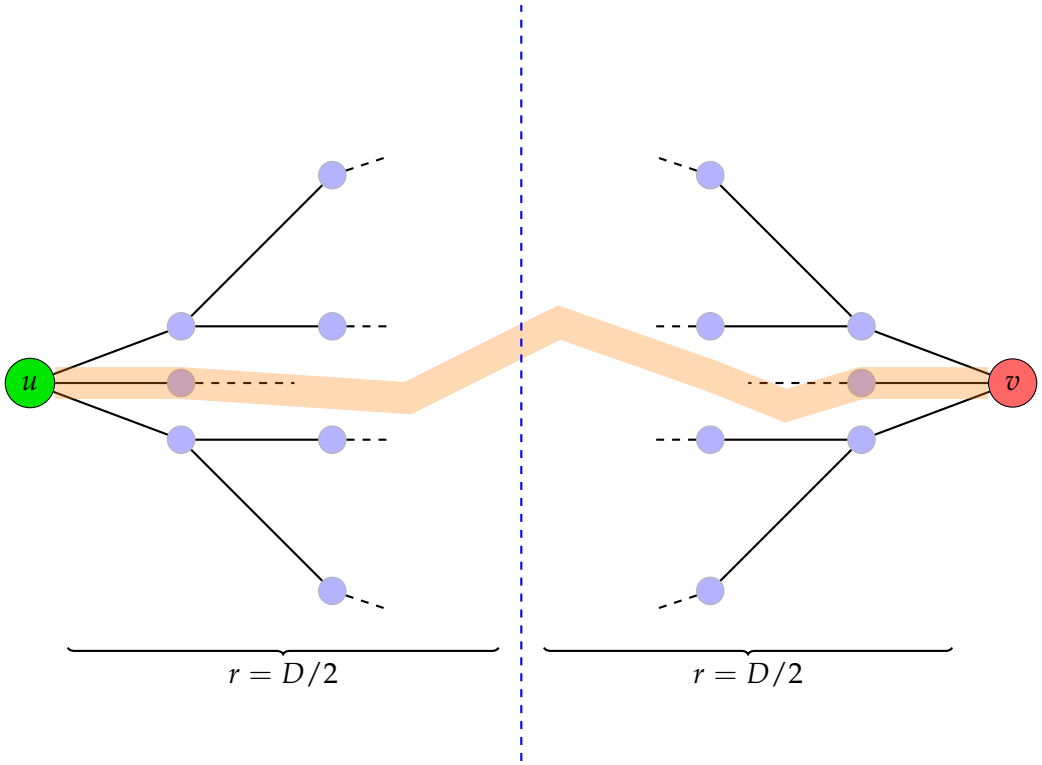


Figure 4: Illustration for the intuition behind Theorem 2.

Theorem 2. *For any vertex-transitive graph G of diameter D and mean distance μ we have $D < 2\mu$.*

Proof. Let u, v be any pair of vertices such that $d(u,v) = D$, and use Lemma 4 to write

$$\mu = \frac{\sum_{x\in V} d(u,x)}{|V|-1} = \frac{\sum_{x\in V} d(v,x)}{|V|-1}, \tag{1}$$

from where

$$\begin{aligned} 2\mu &= \frac{\sum_{x\in V} d(u,x)}{|V|-1} + \frac{\sum_{x\in V} d(v,x)}{|V|-1} \\ &\geq \frac{\sum_{x\in V} d(u,v)}{|V|-1} && \text{(Triangle inequality)} \\ &= \frac{|V|\cdot D}{|V|-1} > D. && \square \end{aligned}$$

We can see that this is tight by considering a cycle on $2n+1$ vertices, where $D = n$ and using Lemma 4,

$$\mu = \frac{1}{2n}\sum_{v\in V} d(u,v) = \frac{1}{2n}\left(2\left(\sum_{i=1}^{n} i\right) + n\right) = \frac{2\cdot\frac{n(n+1)}{2} + n}{2n} = \frac{n}{2} + 1.$$

Similarly, for the hypercube graph Q_n, we have $D = n$ and

$$\mu = \frac{1}{2^n - 1} \sum_{v \in V} d(u, v) = \frac{1}{2^n - 1} \sum_{k=1}^{n} \sum_{\substack{v \in V, \\ d(u,v)=k}} k$$
$$= \frac{1}{2^n - 1} \sum_{k=1}^{n} k \binom{n}{k} = \frac{n 2^{n-1}}{2^n - 1} = \frac{n}{2} + o(1).$$

4 The Demi-god Number for the Rubik's Cube

Theorem 2 allows us to translate upper bounds for the average distance into upper bounds for the diameter. This is particularly useful, as the average distance is easier to certify with high confidence than the diameter. In order to provide an upper bound of the average distance we will follow the following strategy:

- We will sample a large number (500,000) of uniformly random states of the Rubik's cube.
- For each state, we use an efficient solver[4] to obtain an upper bound on the distance, which is certified by the move sequence that the solver outputs.
- We use a simple concentration bound to argue that the empirical average of the distances is a good estimate of the true average distance.

4.1 Randomly sampling cubes

To obtain our estimates on μ_{close} and to determine we need an effective procedure for sampling states of the Rubik's cube uniformly at random. We briefly discuss how to sample a pair of states uniformly at random from the cube.

The *parity of a permutation* is the number of inversions it has modulo 2; i.e., the number of decreasing pairs of increasing entries modulo 2. It is well known, see e.g., [18], that valid cube positions are characterized by just three constraints. Indeed, imagine that we take out all 48 non-center pieces of the Rubik's cube and rearrange them into a new state of the Rubik's cube; the following theorem states when such a rearrangement is a valid state of the cube.

Theorem 3 (Fundamental theorem of Cubology)**.** *A rearrangement of the subcubes is valid if and only if*

- *The permutation of the corners has the same parity as the permutation of the edge.*
- *The number of corners that are twisted clockwise equals the number of corners that are twisted counterclockwise modulo three.*

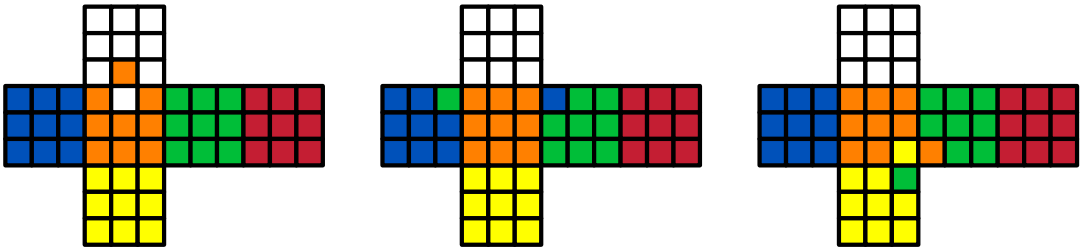
[4]We use https://github.com/efrantar/rob-twophase since it was the fastest solver we could find online.

- *The number of flipped edges is even.*

See [18] for a proof of Theorem 3.

Interestingly, Theorem 3 implies an efficient algorithm for sampling. Consider the following three non-valid operations in the cube:

1. Flip the edge between **F** and **U** (see Figure 5a).
2. Swap the corners at the intersection of **F**, **U**, **L** and **F**, **U**, **R** (see Figure 5b).
3. Turn clockwise the corner at the intersection of **D**, **U**, **R** (see Figure 5c).



(a) Flipping an edge. (b) Swapping two corners. (c) Turning a corner.

Figure 5: Three invalid operations.

Performing these three operations leads to 12 cube configurations (there are 2 choices for flipping the edge, 2 choices for swapping the corners, and 3 choices for turning the corner). Of these 12 configurations, it is easy to see that exactly one of them is valid by Theorem 3, *no matter the state the rest of the cube is in*. Thus, the algorithm that reassembles the cube at random and rejects invalid states by Theorem 3, takes (in expectation), 12 reassembles to uniformly sample a state of the cube. Furthermore, the algorithm that reassembles the cube at random and fixes an invalid state by performing the three operations of Figure 5 takes only 1 re-assembly of the cube to sample uniformly.

Finally, note that, by vertex-transitivity, sampling a random pair in the cube has the same distance distribution as sampling one state of the cube and giving the distance to a *fixed* state. We will use the solved state as the fixed state in our computations and experiments.

4.2 Experimental results

Using the aforementioned methodology, we sampled 500 000 uniformly random states of the Rubik's cube. Out of these, no state required more than 20 moves to be solved, and the empirical mean distance obtained was 18.3189. The process took under 5 hours on a personal computer (MacBook Pro M3, 36 GB of RAM, 16 cores). A histogram is displayed in Figure 6. Our experiments can be found in `https://anonymous.4open.science/r/RubikDemiGodSOSA-E3D5/README`.

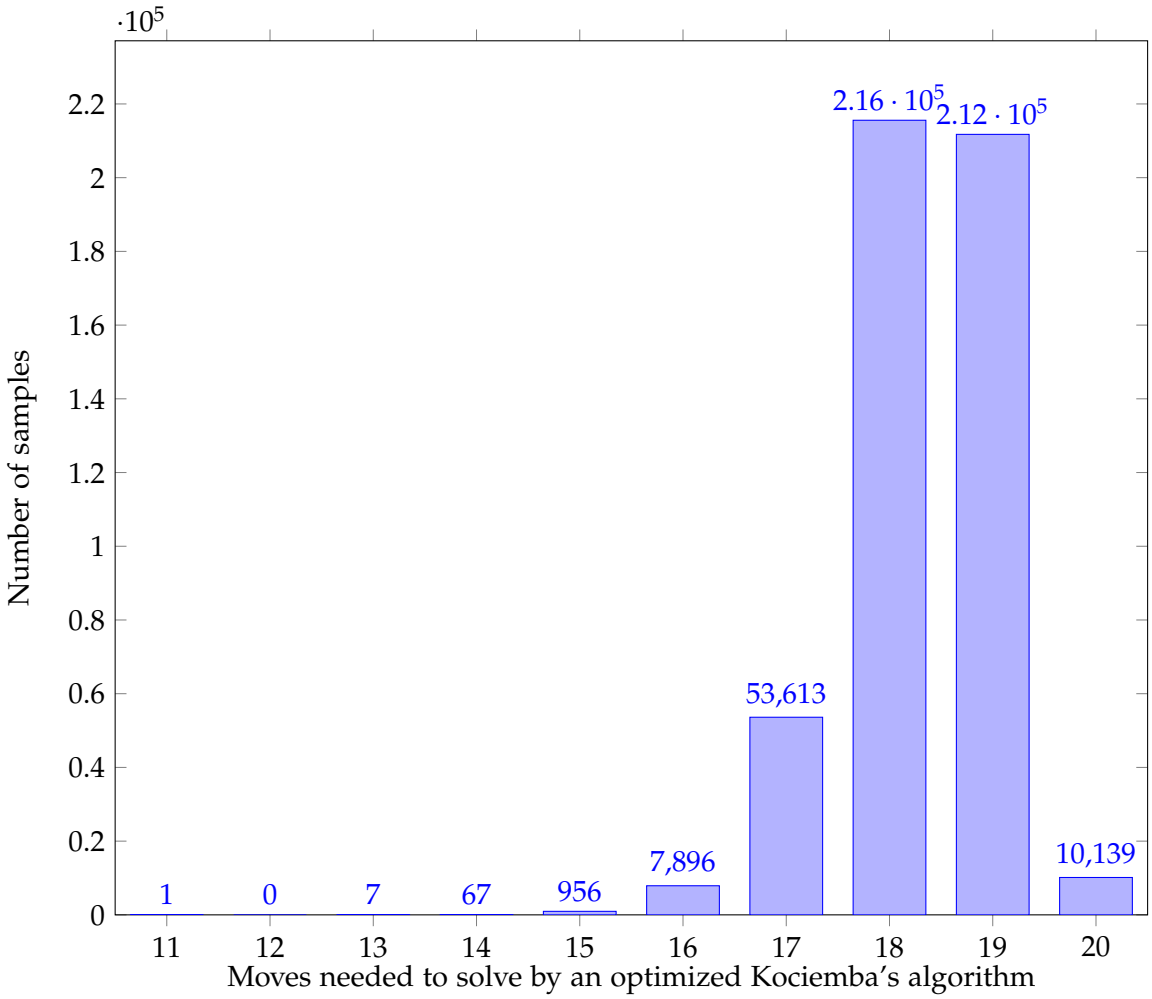


Figure 6: Histogram showing 500 000 samples. This plot aggregates how many moves the samples needed to be solved. No more than 20 moves were needed, and the empirical mean was 18.3189.

4.3 Obtaining the Demi-god's number

We begin by showing that it is highly unlikely that many states of the cube are *"far apart"*. To this end, we say that a pair of states of the cube is *"far apart"* if their distance is 21 or more, otherwise, we say that the states are *"close"*.

We consider the following statement.

"at least 0.03% of the pairs of states of the cube are far apart." (2)

If (2) is true, then by sampling enough pairs of states at random we would expect to observe a pair of states that are far apart. We formalize this idea in the following lemma whose proof is straightforward.

Lemma 5. *If (2) is true, then the probability of sampling s pairs of cube states uniformly at random and observing 0 pair of states that are far apart is at most* $(1 - 0.0003)^s$.

We now consider the mean distance between close vertices, μ_{close}; that is,

$$\mu_{\text{close}} := \sum_{\substack{u,v \in \binom{V}{2} \\ u,v \text{ close}}} d(u,v).$$

Moreover, we consider the empirical mean between close vertices, which we denote by $\widehat{\mu_{\text{close}}}$; that is, we sample pairs at random, compute the distance, discard the results whenever the pairs where far apart, and average the results. A simple concentration result allows us to state that μ_{close} and $\widehat{\mu_{\text{close}}}$ are close to each other with high probability. Indeed, we will use the following version of Hoeffding's inequality.

Lemma 6 (Hoeffding's inequality)**.** *Let $X_1, X_2, \ldots, X_s$ be independent random variables such that $X_i \in [0, C]$.*

$$\mathbb{P}\left(\left|\sum_{i=1}^{s}(X_i - \mathbb{E}[X_i])\right| \geq t \cdot s\right) \leq 2\exp\left(\frac{-2t^2}{C^2 \cdot s}\right).$$

Lemma 7. *Let $\widehat{\mu_{close}}(s)$ be the empirical mean distance over s samples. Then, the probability that $|\widehat{\mu_{close}}(s) - \mu_{close}| \geq 0.1$ is bounded from above by $2\exp(-0.00005 \cdot s)$.*

Proof. For $i \in [s]$, let X_i be the random variable denoting the distance between the i-th pair of states sampled. Note that $|\hat{\mu}_{\text{close}} - \mu_{\text{close}}| \geq 0.1$ if and only if $\left|\sum_{i=1}^{s} X_i - s\mu_{\text{close}}\right| \geq 0.1 \cdot s$. We can now directly apply Hoeffding's inequality to obtain the following.

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^{s} X_i - s\mu_{\text{close}}\right| \geq 0.1 \cdot s\right) &\leq 2\exp\left(\frac{-2(0.1)^2 \cdot s^2}{20^2 \cdot s}\right) \\ &= 2\exp(-0.00005 \cdot s).\end{aligned}$$

□

By Lemma 5, if we assume statement (2), then the probability of observing no pair of states that are far apart when sampling 500,000 pairs of states is at most

$$(1 - 0.0003)^{500\,000} < 7.02 \times 10^{-66}.$$

However, that is exactly what we observed, and these can be easily certified by showing the move sequence that transforms one state into the other (included in our supplementary material). Therefore, we have *overwhelming empirical evidence* that the statement (2) is false. Equivalently, we have overwhelming empirical evidence for:

"at most 0.03% of the pairs of states of the cube are far apart." (3)

Furthermore, the empirical mean observed with 500 000 samples was 18.3189.
Since

$$2\exp(-0.00005 \cdot 500\,000) \approx 2.777 \times 10^{-11},$$

by Lemma 7, we obtain *overwhelming empirical evidence* that

$$\mu_{\text{close}} \leq 18.3189 + 0.1 = 18.4189. \tag{4}$$

Combining facts (3) and (4), we obtain *overwhelming empirical evidence* that

$$\mu \leq \mu_{\text{close}} + 0.03\% \cdot 205 \leq 18.4189 + 0.0615 = 18.4804. \tag{5}$$

Thus, implying our main result.

Theorem 4 (Demi-god's Number (Informal))**.** *There is* overwhelming empirical evidence *that* $D \leq 36$.

Proof. Combine (5) with Theorem 2 to obtain that $D \leq 36.9608$ with overwhelming evidence, and then note that D must be an integer. □

5 Discussion

We have presented a novel approach to bounding the number of moves required to solve any state of the $3 \times 3 \times 3$ Rubik's cube, which relies on two simple aspects of the cube: (i) its vertex-transitivity, and (ii) our ability to efficiently sample uniformly random states from it. Because these two properties extend to a variety of combinatorial puzzles (both to puzzles in the Rubik's family, such as the *Piraminx* or the *Megaminx*, as well as unrelated puzzles like the 15-puzzle on a torus). Moreover, while God's number is known for the $3 \times 3 \times 3$ cube, it remains widely open for larger puzzles [24] (with a gap larger than by a factor of 2 between lower bound and upper bound already for the $5 \times 5 \times 5$ cube), where our approach could be useful provided a good algorithm without requiring theoretical guarantees on it. In terms of related work, So Hirata has recently released two interesting papers concerning the diameter of Rubik's puzzles [12, 13], which attack the problem from different angles; either considering the girth of the cube's graph or using estimations for its branching factor. On a more theoretical line of work, Demaine et al. proved that computing the diameter of an $N \times N \times N$ Rubik's cube is NP-hard [8][5], and that the diameter of the $N \times N \times N$ cube is $\Theta\left(\frac{N^2}{\log N}\right)$ [7].

A particular characteristic of our approach is that our result has an intermediate epistemic status between a theorem and a heuristic, albeit in our opinion much closer to the former. The situation, more in general, is closely related to the question of how much power randomness gives to computation, for which Avi Widgerson recently received a Turing Award [9]; in the context of mathematical results, such a question may be phrased as follows:

Are there properties of finite mathematical objects that can only be certified efficiently to a high degree of confidence by probabilistic algorithms, but that we never can be certain of through a short proof?

[5] More in general, computing the diameter of a Cayley graph is NP-hard given a group presentation [5].

For instance, consider primality testing; whether a given number is prime or not is a fully deterministic fact, in the same way as the average distance of the Rubik's cube graph is. However, in order to practically obtain knowledge of such deterministic facts, we leverage the computational benefit of randomness, which allows us (at least in current practice), to determine facts that otherwise would be out of reach. The cost, however, is the possibility of error in the associated randomized algorithms, which forbids us from claiming to have definite proofs of the facts of interest. As usual, we can get such probability of error to be as small as we deem necessary for convincing ourselves, at a modest computational price, while keeping the curse of never reaching 100% confidence. An interesting counterpoint is to discuss whether traditional proofs equal certainty, as it is not evident that when reading traditional proofs we can reliably reach 100% of confidence either. We might claim that for simple proofs like the irrationality of $\sqrt{2}$, the elementary proof of Theorem 2 in this paper, or even the Central Limit Theorem. However, proofs that span dozens or even hundreds of pages, covering a multitude of cases, and including non-trivial calculations, are much more delicate from a trust perspective. For instance, the proof of Kepler's conjecture by Thomas Hales took years before reviewers, from the prestigious Annals of Mathematics, accepted the paper while saying they were only "99% sure of its correctness" [10, 16]. For a more general discussion of the impact of computation in modern mathematics, and how our understanding of "proofs" can be affected by computation, we refer the interested reader to the work of Avigad [1, 2].

Going back to our case, we have shown that, if one assumes momentarily that the diameter of the Rubik's cube graph is larger than 36, then observing an emipirical mean distance of around 18.3 over 500 000 samples, none of which required more than 20 moves, has probability under 10^{-10}. We encourage the readers to reproduce this computation by themselves, which should take less than a day in any modern computer. We believe this makes an extremely compelling case for the diameter of the Rubik's cube graph being at most 36 while using a fraction of the computation required by previous approaches. Moreover, we hope that this same line of attack can be useful for analyzing other puzzles or graphs.

Acknowledgments. The second author thanks Ilan Newman for a discussion that helped simplify the proof of Theorem 3, and Jeremy Avigad for references on "plausibility". We thank Elias Frantar for making his efficient Rubik's cube solver publicly available, which facilitated this project.

References

[1] Jeremy Avigad. Varieties of mathematical understanding. *Bulletin of the American Mathematical Society*, 59(1):99–117, February 2021.

[2] Jeremy Avigad. Varieties of mathematical understanding. https://arxiv.org/abs/2310.20100v1, October 2023.

[3] Christoph Bandelow. *Inside Rubik's Cube and Beyond*. Birkhäuser Boston, 1982.

[4] Jonathan Borwein and David Bailey. *Mathematics by Experiment: Plausible Reasoning in the 21st Century*. A K Peters/CRC Press, New York, 2 edition, May 2011.

[5] Jin-Yi Cai, George Havas, Bernard Mans, Ajay Nerurkar, Jean-Pierre Seifert, and Igor Shparlinski. On Routing in Circulant Graphs. In Takano Asano, Hideki Imai, D. T. Lee, Shin-ichi Nakano, and Takeshi Tokuyama, editors, *Computing and Combinatorics*, pages 360–369, Berlin, Heidelberg, 1999. Springer.

[6] JJ Chen. Group theory and the rubik's cube, 2004.

[7] Erik D. Demaine, Martin L. Demaine, Sarah Eisenstat, Anna Lubiw, and Andrew Winslow. Algorithms for solving rubik's cubes. In Camil Demetrescu and Magnús M. Halldórsson, editors, *Algorithms - ESA 2011 - 19th Annual European Symposium, Saarbrücken, Germany, September 5-9, 2011. Proceedings*, volume 6942 of *Lecture Notes in Computer Science*, pages 689–700. Springer, 2011.

[8] Erik D. Demaine, Sarah Eisenstat, and Mikhail Rudoy. Solving the rubik's cube optimally is np-complete. In Rolf Niedermeier and Brigitte Vallée, editors, *35th Symposium on Theoretical Aspects of Computer Science, STACS 2018, February 28 to March 3, 2018, Caen, France*, volume 96 of *LIPIcs*, pages 24:1–24:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[9] Association for Computing Machinery. Acm a.m. turing award honors avi wigderson for foundational contributions to the theory of computation, 2023. [Accessed 06-08-2024].

[10] Thomas Hales. A proof of the kepler conjecture. *Annals of Mathematics*, 162(3):1065–1185, November 2005.

[11] Mark Herman and Jonathan Pakianathan. On the distribution of distances in homogeneous compact metric spaces. *Topology and its Applications*, 193:97–99, 2015.

[12] So Hirata. Graph-theoretical estimates of the diameters of the rubik's cube groups, 2024.

[13] So Hirata. Probabilistic estimates of the diameters of the rubik's cube groups, 2024.

[14] Nathan Jacobson. *Basic Algebra I*. Dover Books on Mathematics. Dover Publications, Mineola, NY, 2 edition, June 2009.

[15] JPerm. Learn How to Solve a Rubik's Cube in 10 Minutes (Beginner Tutorial). https://www.youtube.com/watch?v=7Ron6MN45LY, 2019. [Accessed 03-08-2024].

[16] Roxanne Khamsi. Mathematical proofs getting harder to verify. https://www.newscientist.com/article/dn8743-mathematical-proofs-getting-harder-to-verify/, 2006. [Accessed 06-08-2024].

[17] Herbert Kociemba. Close to god's algorithm. In *Cubism for Fun*, pages 10–13, 1992.

[18] James T. Mulholland. Permutations puzzles - rubik's cube. https://www.sfu.ca/~jtmulhol/math302/puzzles-rc-cubology.html#:~:text=The%20Fundamental%20Theorem%20of%20Cubology%20(plain%20language%20form)&text=The%20number%20of%20corners%20that,of%20flipped%20edges%20is%20even. [Accessed 08-08-2024].

[19] Ilan Newman and Yuri Rabinovich. Hard Metrics from Cayley Graphs of Abelian Groups. In Wolfgang Thomas and Pascal Weil, editors, *STACS 2007*, pages 157–162, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[20] G. Polya. *Mathematics and Plausible Reasoning, Volume 1: Induction and Analogy in Mathematics*. Princeton University Press, 1954.

[21] Tomas Rokicki, John Dethridge, Herbert Kociemba, and Morley Davidson. God's Number is 20. https://www.cube20.org/. [Accessed 08-08-2024].

[22] Tomas Rokicki, John Dethridge, Herbert Kociemba, and Morley Davidson. God's Number is 26 in the Quarter Turn Metric. https://cube20.org/qtm/. [Accessed 08-08-2024].

[23] Tomas Rokicki, Herbert Kociemba, Morley Davidson, and John Dethridge. The Diameter of the Rubik's Cube Group Is Twenty. *SIAM Review*, 56(4):645–670, January 2014.

[24] Speedsolving.com Wiki. God's Algorithm - Optimal solutions of the Rubik's Cube. https://www.speedsolving.com/wiki/index.php/God%27s_Algorithm#Table_of_God.27s_Numbers, 2024. [Accessed 06-08-2024].

[25] Douglas B. West. Problems in Graph Theory and Combinatorics. http://dwest.web.illinois.edu/openp/. [Accessed 02-08-2024].

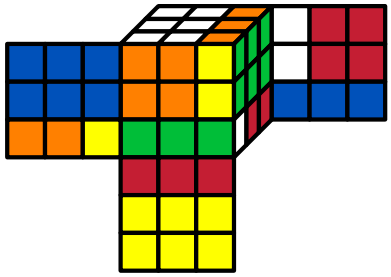
[26] Baoyindureng Wu, Guojie Liu, Xinhui An, Guiying Yan, and Xiaoping Liu. A conjecture on average distance and diameter of a graph. *Discrete Mathematics, Algorithms and Applications*, 03(03):337–342, September 2011.

A The Beginner's Method and the "Human's Number"

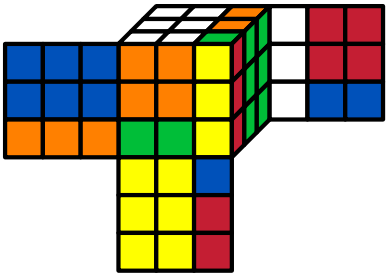
In order to make this article self-contained, we provide a brief overview of the *"beginner's method"* to solve the Rubik's Cube[6], from which we have the following:

Lemma 8 (Human's Number)**.** *Any position of the Rubik's cube can be solved in at most 205 moves.*

[6]We encourage, nonetheless, the interested reader to look into the many YouTube videos (e.g., [15]) that guide the process.



(a) Result of **RD'** from the solved state.



(b) Result of **D'R** from the solved state.

Figure 7: Illustration of the non-commutativity of the Rubik's cube.

In a nutshell, the beginner's method consists of solving the Rubik's Cube by *"layers"*, as opposed to by faces. Before we begin with its exposition, however, it is worth establishing some notation for the different Rubik's cube moves. We will use "Singmaster" notation, credited to British mathematician David Singmaster. To specify a turn on a face, we use the first letter of the face's name: **R** for the right face, **L** for the left face, **U** for the upper face, **D** for the down face, **F** for the front face, and **B** for the back face. If the face is to be rotated by 90° clockwise, we add no suffix, e.g., **R** means a clockwise rotation by 90° of the right face. For a counterclockwise rotation by 90°, we add a prime symbol, e.g., **D'** means a counterclockwise rotation by 90° of the down face. For a 180° rotation, we add a 2 after the letter, e.g., **F**2 means a 180° rotation of the front face. A proper formalization of what a *"move"* actually is can be found in Section 2, where we view the Rubik's cube as a group. As an example to check our understanding of the notation, the non-commutativity of the Rubik's cube group is evidenced in Figure 7.

Step 1: The white cross The first step consists of creating a *"cross"* on one face of the cube, which we will assume to be white without loss of generality. To achieve this, one must move every white "edge" (i.e., a piece with two colors) to the correct position. That is, e.g., the white-orange edge must be placed so that its white sticker is adjacent to the white center and its green sticker is adjacent to the orange center. We illustrate the result of this step in Figure 8, and a conservative bound is that this can always be achieved in 20 moves, as each of the 4 white edges to place can always be placed in at most 5 moves or fewer. This step can be done intuitively, that is, without memorizing any particular algorithm[7].

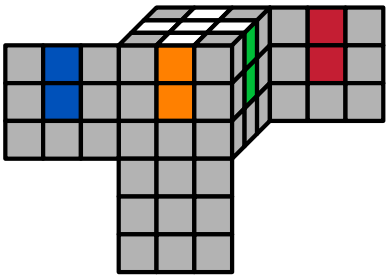


Figure 8: The white cross is solved.

[7] In the Rubik's cube literature, a move sequence with a concrete purpose (e.g., permuting 3 corner pieces) is traditionally called an "algorithm".

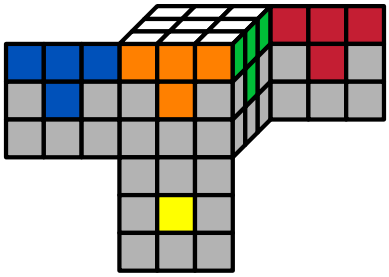


Figure 9: The white corners are solved, and thus the first layer is completed.

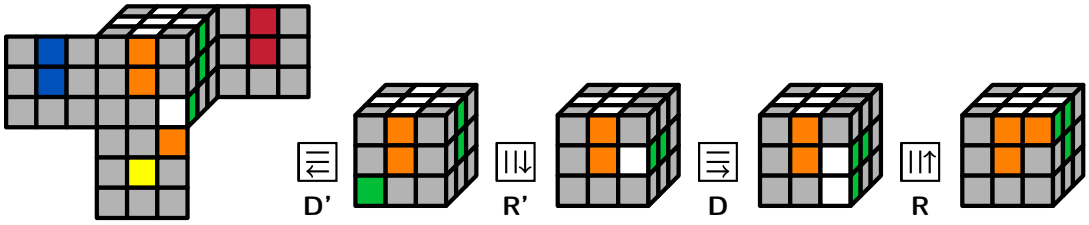


Figure 10: Illustration of one of the cases for placing a corner in the first layer (Step 2 of the beginner's method), through the move sequence **D'R'DR**.

Step 2: White corners The second step consists of placing the white corners in their correct position, one by one. To place a white corner, one can first bring it to the opposite layer (i.e., the bottom layer, whose center is yellow), and then proceed according to a handful of cases, one of which is illustrated in Figure 10. The result of this step is illustrated in Figure 9. Conservatively, this step can always be achieved in 15 moves per corner, and 60 in total. This accounts for the cases when a white corner is in the correct location but oriented incorrectly, in which case a non-white corner can be placed in that spot, thus allowing the white corner to be placed in the correct orientation afterward.

Step 3: Edges of the second layer The third step consists of placing the edges of the second layer in their final position, as illustrated in Figure 11. For instance, the orange-green edge must be placed so that its orange sticker is adjacent to the orange center and its green sticker is adjacent to the green center. The main algorithm to solve this step is illustrated in Figure 12. A conservative bound, again due to cases in which a misoriented edge must be first replaced before placing it in the correct orientation, is 20 moves per edge.

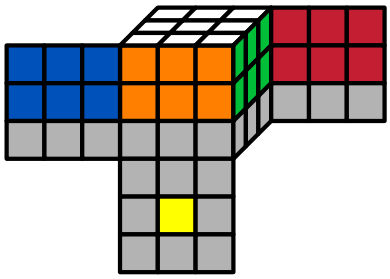


Figure 11: The first two layers are solved.

Step 4: The yellow cross We now turn our attention to the yellow face. The goal of this step is to solve the orientation of the yellow cross,

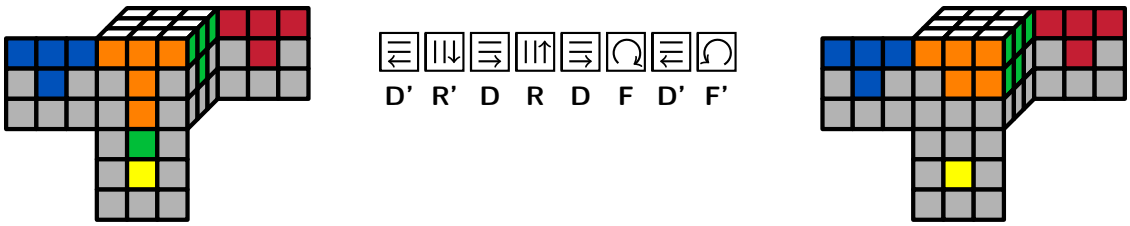
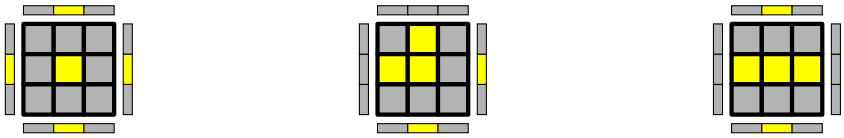


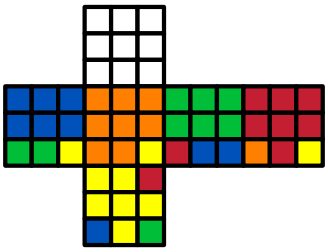
Figure 12: Illustration of the algorithm to solve the edges of the second layer (Step 3 of the beginner's method).



(a) No edges oriented correctly.

(b) Non-opposite edges oriented correctly.

(c) Opposite edges oriented correctly.



(d) A possible state after solving the yellow cross.

Figure 13: Illustration of Step 4 of the beginner's method.

that is, to make all yellow edges have their yellow sticker adjacent to the yellow center, as depicted in Figure 13d. We may face three different scenarios in this step (if it is not already solved), as illustrated in Figure 13. These can all be solved by the same algorithm, potentially repeated according to which of the three non-solved cases we encounter. The algorithm is simply: **FRUR'U'F'**. Applying it from case 13a leads to case 13b, and applying it again leads to case 13c, from where a last application solves the yellow cross. That way, we need at most 3 applications, leading to a conservative bound of 18 moves for this step.

Step 5: Permuting yellow edges Now, we permute the yellow edges so that each of them gets to its desired position. This step can be solved by repeated application of a single algorithm, that induces a 3-cycle of the yellow edges, as illustrated in Figures 14 and 15b. As this algorithm is applied at most 3 times, we have a conservative bound of 21 moves for this step.

Step 6: Permuting yellow corners This step is analogous to the previous one but over the corners; we permute the yellow corners so that each of them gets to its desired position. This step can also be solved by repeated application of a single algorithm, that

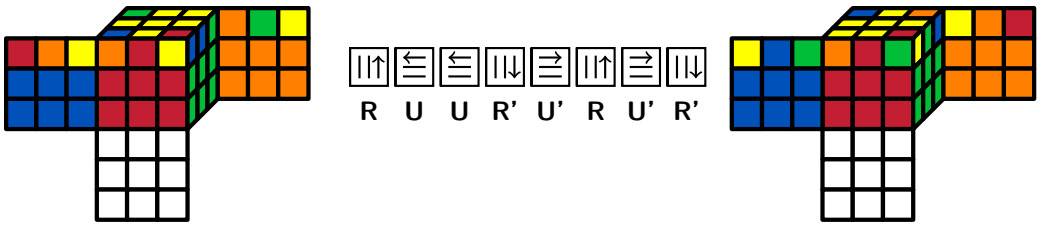
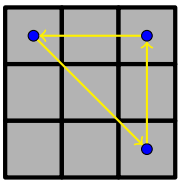
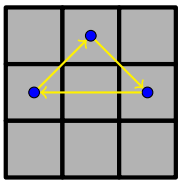


Figure 14: Illustration of a case for Step 5 of the beginner's method.



(a) 3-cycle permutation of corners induced by **RU'L'UR'U'LU**.



(b) 3-cycle permutation of edges induced by **RU2R'U'RU'R'**.

Figure 15: Illustration of the 3-cycle algorithms for permuting yellow corners and edges, corresponding to Steps 5 and 6 of the beginner's method.

induces a 3-cycle of the yellow corners, as illustrated in Figures 15a and 16. This algorithm is applied at most 3 times, leading to a conservative bound of 24 moves for this step.

Step 7: Orienting yellow corners The last step consists of orienting the yellow corners, which again can be achieved by repeated applications of a single algorithm that changes the orientation of two adjacent corners (illustrated in Figure 17):

RU2R'U'RU'R'L'U2LUL'UL.

This algorithm needs to be applied at most 3 times, leading to a conservative bound of 42 moves for this step.

The following table summarizes the "proof" of Lemma 8:

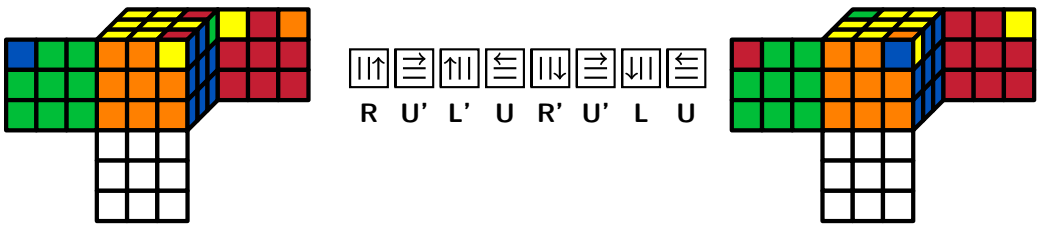


Figure 16: Illustration of a case for Step 6 of the beginner's method.

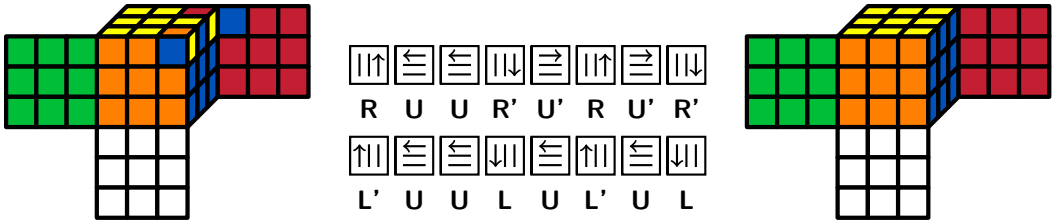


Figure 17: Illustration of a case for the final step of the beginner's method, using the move sequence: **RU2R'U'RU'R'L'U2LUL'UL**.

Step	Moves
White cross	20
White corners	60
Edges of the second layer	80
Yellow cross	18
Permuting yellow edges	21
Permuting yellow corners	24
Orienting yellow corners	42
Total	205