

Title: Cracking the Collatz Code: A Journey from Conjecture to Certainty through Infinity

Author: Eric Lough

Email: neverquietmind@gmail.com

Abstract:

The Collatz conjecture, a puzzle that has intrigued mathematicians for decades, has met its resolution through a simple combination of trial and error combined with a spreadsheet. While the primary goal was to definitively prove the conjecture, the exploration revealed unexpected insights into the underlying patterns of the Collatz sequence. This study introduces a novel approach to organizing and analyzing Collatz data, laying bare the formulas governing the sequence's behavior. The results present a conclusive proof of the Collatz conjecture and unveil a spectrum of sets and formulas reaching to infinity, each contributing to a deeper understanding of its workings. Comprehensive instructions are provided, encouraging collaboration and furthering the collective understanding of mathematical sequences.

Methodology:

When I began seriously working on understanding Collatz, the first thing I did was to carefully examine the fundamental mathematical operations involved. The sequence of result numbers operates through two primary operations. If the number is odd, then multiply by 3, and add 1. If the number is even, divide by 2. From here I decided that  $3n+1$  is up, and  $n/2$  is down.

My initial exploration started with a spreadsheet. I created a column of seed integers between 1 and 100. To the left of that column, I entered in a function that would read the integer in the seed column and perform  $3n+1$  if the integer in the seed column for that row was odd. To the right of the seed column, I entered in a function that would read the integer in the seed column, and perform  $n/2$  if the integer was even.

As I continued my exploration in this manner, I noticed that between 4 and 10, was a difference of 6 numbers. This meant that  $3n+1$  performed on any odd number would only hit every 6<sup>th</sup> even number. So I modified the evens to show +1 if the seed was odd, and the odds to show +6 if the seed was even.

Odds	Seed	Evens
4	1	
+6	2	1
10	3	+1
+6	4	2

The next thing I did was to copy the formulas in the Odds and Evens Column into an empty area in the spreadsheet, leaving a central column empty starting at L1 for typing in a seed manually. Within this central column I placed this formula in L2 '=IF(ISODD(L1); K1; M1)' where K1 had '=IF(ISODD(L1); L1\*3+1; "+6")' and M1 had '=IF(ISEVEN(L1); L1/2; "+1")' Dragging each formula down to row 500, while leaving L1 empty to type in seed integers.

As I carefully observed the results from each integer between 3 and 100 in my new column, I started to notice a pattern, that led me to set of powers of 2. I was having some difficulty in correlating my results, so I then decided to write my own program using Godot Game engine, to calculate the integers in batches, starting with 3 to 100, then increasing and running 100 integers at a time for each time I clicked the generate button. One of the difficulties I was running into was getting openOffice spreadsheet to indicate the prime numbers, this was solved by writing my own program. While the program is rudimentary, simply using debug prints to get the data I wanted, it was effective as I was able to start noticing more patterns between the numbers. I had thought that perhaps prime numbers had something to do with the randomness, but further exploration proved that thought incorrect.

I noticed early on while running my program that many of the numbers tested would leave on their final descent at 16, so I adjusted the program itself to hold a global array containing all of the powers of 2, from 1 up to 8388608, the first 24 powers of 2. I also created a global sister array that held 24 0 placeholder values. I modified the program itself to scan the powers of 2 array, and to note which integer within the powers of 2 was reached, where the program would count the number of times the current 100 integers would use which power of 2 to make their way to 1. Later I also added a third array of 24 placeholder 0's to see how many perfect sets I could find. A perfect set was defined as every integer tested in that set of 100 integers would all use the same power of 2 integer to leave toward 1. The results of my findings led me to just delete the perfect sets, because 16 was the most popular number. So popular that in the first 300K integers I tested with my program 93% exited at 16. I found this result challenging because the higher I went the more numbers that would leave using 16, on occasion, they'd also sometimes exit on 64, and less frequently on 256 and 1024. These are all powers of 4.

At this point, I felt I had enough data from my program to go back to exploring the results with open office spreadsheets. One of the things I had noticed before was that  $3n+1$  reached even integers, with a gap of 6 integers between. So performing  $3n+1$  on 3 resulted in 10, and on 5 resulted in 16. This meant that 12, and 14 are unreachable by  $3n+1$ . So I created a new spreadsheet called Holes, intended to explore the holes between the even integers you can reach with  $3n+1$ .

I started this sheet with all of my seed integers between 1 and 200 split between 2 columns, one for odds, and one for evens. To the left of the odds column I placed 3 more columns, the first one to the left of odds is result up, which shows the result of  $3n+1$  performed on the odd numbers in that column and row. To the left of that was Miss 1, and to the left of that was Miss 2. to the right of the evens was a column named result down, showing  $n/2$ .

Miss 2	Miss 1	Result Up	Odds	Evens	Result Down
8	6	4	1	2	1
14	12	10	3	4	2
20	18	16	5	6	3
26	24	22	7	8	4
32	30	28	9	10	5

Turns out performing  $3n+1$  on 1 is beneficial for finding even integer holes. An additional thing I started doing here was color coding inside my evens column, specifically powers of 2, powers of 4, and shading numbers that are evenly divisible by integers that are in the powers of 2, but are not powers of 2 themselves. The bigger the gap between the powers of 2 in the evens, the more even numbers that are divisible by powers of 2, and by more of them. For example, between 8 and 16 is 12, evenly divisible by 2 and 4. Between 16 and 32, we have 20 and 28, evenly divisible by 2 and 4, and 24 evenly divisible by 2, 4 and 8. Between 32 and 64, we have integers evenly divisible by 2 and 4, 3 integers evenly divisible by 8, and 1 integer evenly divisible by 16. This pattern continues for each gap between the powers of 2, meaning as the gaps increase in distance, the numbers that are evenly divisible by powers of two increase, along with the number of integers inside of the powers of 2 set that can evenly divide those numbers increase.

After this I went back to my first spreadsheet, with holes opened next to it, I created a new sheet inside of the first and added a column of seed integers from 1 to 5000, and then a column of powers of 2, which I stopped at  $2^{70}$ , wholly unnecessary looking back. To the right of that I started a set defined as all the even integers you can reach with  $3n+1$ , to the right of that a subset of the  $3n+1$  where performing  $n/2$  on the result of  $3n+1$  would give you another even result, a subset to the right of that where performing  $n/2$  on the result of  $3n+1$  would give you an odd result. All told, using the data sorted into the previous sets I separated and identified 11 total sets you could define with rules for how to enter and leave each set.

Unfortunately this initial layout was flawed and my set definitions were also flawed, but as I worked through them I added a conditional formatting to seed column to bold and color all the seed integers whenever there was a duplicate integer matching the seed column integers. It wasn't long before

every integer was bolded and colored in my seed column up to 5000, but there were gaps when I extended past 5000. So I felt I was missing something. I went back to the first sheet and spent a few days comparing Collatz results where there were a lot of steps.

I was pondering the formulas for Collatz, multiply by 3, divide by 2. and I had the thought of attempting to simplify Collatz. When you perform  $3n+1$ , you make an even number which means you immediately follow it with  $n/2$ . When you do this you wind up with one of three types of even integers. The first type of even integer is one that divides into an odd, the second type is one where you have to perform  $n/2$  at least twice before you reach an odd integer, and the third kind seemed to be an even integer where you perform  $n/2$  at least 4 times before you reach an odd integer.

For example, take 3, perform  $3n+1$  to get 10, then  $n/2$  and you get 5. We went up but came back down. We came down not as far as we went up, but we still came down. If we divide 5 by 3, we get 1.66. I didn't feel this was accurate enough. I randomly decided to try to use 3 and 2, so when we divide 3 by 2 we get 1.5.  $3*1.5 = 4.5$ , round that and we get 5. We just skipped 10, but made it to 5.

I bolded the primes to demonstrate we don't miss any of the primes.

27 ran with  $3n+1$ . [27, 82, **41**, 124, 62, **31**, 94, **47**, 142, **71**, 214, **107**, 322, 161, 484, 242, 121, 364, 182, 91, 274, **137**, 412, 206, **103**, 310, 155, 466, **233**, 700, 350, 175, 526, **263**, 790, 395, 1186, **593**, 1780, 890, 455, 1336, 668, 334, **167**, 502, **251**, 754, 377, 1132, 566, **283**, 850, 425, 1276, 638, 319, 958, **479**, 1438, **719**, 2158, 1079, 3238, **1619**, 4858, 2429, 7288, 3644, 1822, **911**, 2734, **1367**, 4102, 2051, 6154, 3077, 9232, 4616, 2308, 1154, **577**, 1732, 866, **433**, 1300, 650, 325, 976, 488, 244, 122, **61**, 184, 92, 46, **23**, 70, 35, 106, **53**, 160, 80, 40, 20, 10, **5**, 16, 8, 4, **2**, 1]

27 ran with  $\text{round}(n*1.5)$  [27, **41**, 62, **31**, **47**, **71**, **107**, 161, 242, 121, 182, 91, **137**, 206, **103**, 155, **233**, 350, 175, **263**, 395, **593**, 890, 445, 668, 334, **167**, **251**, 377, 566, **283**, 425, 638, 319, **479**, **719**, 1079, **1619**, 2429, 3644, 1822, **911**, **1367**, 2051, 3077, 4616, 2308, 1154, **577**, 866, **433**, 650, 325, 488, 244, 122, **61**, 92, 46, 23, 35, **53**, 80, 40, 20, 10, **5**, 8, 4, **2**, 1]

Using  $\text{round}(n*1.5)$  on 27 is 37% more efficient at reaching 1 than the standard  $3n+1$ .  $3n+1$  uses 111 steps to reach 1, performing  $3n+1$  41 times and  $n/2$  70 times, while  $\text{round}(n*1.5)$  uses 70 steps to reach 1 performing  $\text{round}(n*1.5)$  41 times, and  $n/2$  29 times. As you should also notice, the new peak is 4616. What happens when we run  $\text{round}(n*1.5)$  instead of  $3n+1$ , is we wind up skipping the first even result, if there are multiple even results in that chain, we will keep skipping the first even integer in our results, but we don't miss any of the odds.

If you think of Collatz like the game snakes and ladders, performing  $3n+1$  is like climbing a ladder, while  $n/2$  is like falling down a snake. If the result of the next  $n/2$  after performing  $3n+1$  is an odd integer,  $\text{round}(n*1.5)$  simply places a staircase to the next odd integer, and places fireman pole between the current odd integer and the even integer after the next even integer. Should one want to use Collatz as a prime number sieve, this modified formula will speed things along. It is entirely possible there are other shortcuts in Collatz, possibly ones that are more efficient, however, that wasn't my focus, and even though we're performing  $n/2$  less than half as often and still reaching 1, I didn't think this was enough proof to submit, so I kept digging, and looking for sets.

I kept looking at the formulas, and trying to define sets of integers, the first set I initially thought should be all powers of 2, but the first set of integers you should define when working Collatz is the seed integer set. I stumbled upon this realization when I started really looking at the even integer results of  $3n+1$ . Consider this set to be [10, 16, 22, 28, 34, 40, 46, 52, 58, 64,...] simply add 6 to the last integer in the set, to find the odd that got you there subtract 1 and divide by 3. Every even integer in this set has 2 ways in, and one way out. One of the ways in is by performing  $3n+1$ . The other way in is to perform  $n/2$  on the even integer that is twice the value. And the only way out is  $n/2$ . Every time one of these integers appears in any of the visualization trees I've found for Collatz, these are the branches. 2 ways in, 1 way out.

If we look at all the evens we can't reach with  $3n+1$ , those are all one way in, one way out, via  $n/2$ . In addition, all the odd numbers have 1 entrance and one exit, you reach them by  $n/2$  and leave them with  $3n+1$ . The only caveat to these rules is if you used the integer as the seed to start the operations.

Unfortunately things aren't as simple as taking all the even numbers we find we can't reach with  $3n+1$  and putting them into their own set. I fiddled with this a lot, and found sets that had various rules, but were tricky to define how a number wound up in this set. Eventually I discovered that the way most people present the visualization of how the Collatz sequence works is flawed.

If you want to properly visualize Collatz and all of the many nuances and interactions between the sets, you have to lay it out in a table. Everything you ever wanted to know about Collatz reveals itself when laid out in the proper format.

After I laid it out properly and started examining it, I realized there were only 2 integers out of every 8 that traveled upward, and 6 travel downwards. It was fairly simple to figure out the rules for 2 steps up and 3 steps down but then things got weird, and I could see a pattern but I couldn't initially figure out how to identify it. I wound up writing up the first version of my proof before I finished identifying all the patterns. As I kept trying to find the right words to explain why Collatz gets weird in certain places, I kept tossing ideas at it to try to solve it once and for all. It was just that last little bit that was confusing.

I had initially identified 3 distinct formulas for working Collatz, but as I continued to try to make the weird numbers work, I realized that the number of formulas you actually need to work it without using  $3n+1$  and  $n/2$ , where you can count the steps, and figure out how you enter and leave each set, without brute forcing your way through it, are infinite. Once I realized the pattern and how to find it, I had to work on how to construct the formulas so you can use your starting point, and get an ending point, along with the number of steps. In my results section below, I walk you through how to identify where you need a new formula, and how to construct it so you can keep chasing whatever number you want to find.

The formulas for solving Collatz are infinite because the powers of 2 are infinite.

Results:

Here is the best way to visualize Collatz.

Create your seed integer line. The 0 is used for the step counter.

0	1	2	3	4	5	6	7	8	9	10	11	12	13
---	---	---	---	---	---	---	---	---	---	----	----	----	----

For every space above the seed integer line, we proceed to multiply by 2 to infinity, and we fill in the row underneath our seed integers with  $n/2$  or  $3n+1$ , repeating whichever is applicable until we either reach 1, or another odd integer.

	*2	*2	*2	*2	*2	*2	*2	*2	*2	*2	*2	*2	^2
	inf												
	32	64	96	128	160	192	224	256	288	320	356	384	416
	16	32	48	64	80	96	112	128	144	160	176	192	208
	8	16	24	32	40	48	56	64	72	80	88	96	104
	4	8	12	16	20	24	28	32	36	40	44	48	52
Steps	2	4	6	8	10	12	14	16	18	20	22	24	26
0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	1	10	2	16	3	22	4	28	5	34	6	40
2	4		5	1	8		11	2	14		17	3	20
3	8				4			1	7				10
4	16				2								5
5	32				1								

Each seed integer on the seed integer line is considered a full set. The set begins at the seed integer, and travels downward to the next odd integer. For each set you enter, you travel the indicated number of steps. The powers of two extending below 1, are to indicate how far to the right you'll find the next integer that uses the same rules as the ones you can currently view.

Immediately above our seed integer line, we are counting by 2's all the way across. Above that by 4's, above that by 8's, then 16's, 32's, 64's, etc all the way to infinity. Any even integer you'll ever need exists above the seed integer line.

The far left column that descends to infinity from 0 is your step count.

Notice the seed column starts at 0, and goes all the way to infinity to the right. This table view stops at 13.

From 1 we go up by powers of 2 to infinity, and we go down by powers of 2 to infinity as well.

When we fill in integers below the seed integer line we process Collatz as normal, but as soon as we reach 1, or another odd integer we stop processing.

Everything you need to unravel Collatz completely is either in this table right now, or is hinted at in this table. There's still a bit of leg work left to improve the formulas, which will require a lot of time spent with a spreadsheet pouring through Collatz results. But I'll explain the formulas I have, and how to find the rest. When I first examined this table it didn't have the extra column to the left, that was a recent discovery, and the extra column at the left means Collatz has no mysteries left.

As I mentioned in my Methodology, I discovered that you can use  $\text{round}(n*1.5)$  as a substitute for  $3n+1$ . In the example where I processed 27 using both  $3n+1$ , and the shortcut  $\text{round}(n*1.5)$ , both processes went up 41 times. However the shortcut only needed  $n/2$  29 times instead of the normal 70 times. This is the first clue that led me to arranging Collatz in this way. What's important in the Collatz Sequence is the odd integers, not the evens. So we start with our seed integer line, and we stop when we come across an odd integer.

Many people got hung up on the powers of 2 set, and it is instrumental in solving Collatz, but maybe not the way you thought. First, even though powers of 2 are an immediate exit from Collatz to 1, very very few numbers actually enter the powers of 2 set to exit Collatz except through very specific locations, like 5 to 16. For example, 32 is not an integer you can reach except as you pass it from 64 coming down. There are a lot of powers of 2 that aren't actually reachable from any seed unless that integer is used as a seed, or you pass through it from above.

For this explanation, it would be a very good idea if you were to continue filling out a spreadsheet with Collatz data, just keep adding to your seed integer line, fill out the spots above the seed integer line by doubling whatever the integer is a few times, and then perform  $3n+1$  or  $n/2$  below until you reach the next odd number. I've found a combination function like `'=IF(ISODD(C14);C14*3+1;C14/2)` ' is super handy for filling out the data immediately underneath the integer line, and I use `'=IF(C15 = " "; " ";IF(ISEVEN(C15); C15/2; " "))'` for all the lines following so as soon as I hit another odd, the next lines are blank. It helps when navigating the results and is extremely handy when building formulas that make Collatz spill all of its secrets.

Let us start with 3 using our table above, we just pop right down to 5 and then look to the left and we see we have 2 steps from 3 to 5, then we go to 5 and at the bottom we see 1, look at our step count and we get 5 steps to 1 from 5,  $5+2 = 7$ , It takes 3 7 steps to reach 1. But that's not why that column for counting steps is there. That column and the increasing values of powers of 2 are there for a completely different reason.

Let us start with 3 again. 3 goes to 5, it takes 2 steps. 3 goes UP. 7 goes to 11, it takes 2 steps, 7 goes UP. 11 goes to 17, it takes 2 steps. 11 goes UP. If we take the number we went to, and subtract the number we came from, we get an interesting result.  $5-3 = 2$ ,  $11-7=4$ ,  $17-11 = 6$ .

Let us look at our steps column again. Immediately to the right of 2 steps is 4. We need that 4 in our formula to calculate what odd integer we reach when we go UP.

UP =  $(x-3)/4 + 1$ . As long as UP is an integer and not a float, the odd number we plug into x, is every other odd number starting with 3, and we go up by  $x + UP*2$ .

If x is 3, UP =  $3-3/4 + 1$ .  $3-3 = 0 / 4 = 0 + 1 = 1 * 2 = 2$ .  $3+2 = 5$ . Add 2 steps.

If x is 7, UP =  $7-3/4 + 1$ .  $7-3=4/4 = 1+1=2*2= 4$ .  $7+4 = 11$ . Add 2 steps.

Next let us look at 5, 9, 13, etc... all of these go down. We can't actually process 5, 9 and 13 with the same formula, but they all do the same thing, the trick is the number of steps we go down is different for each one.

DOWN =  $(x-1)/4$ . If DOWN is an integer and not a float, well, this is where the fun starts.

We have all kinds of odd integers that go down, anywhere from 3 steps to an infinite number of steps back down, and they happen every other integer starting at 5.

The question becomes what is x evenly divisible by? Perhaps the most efficient way to figure this out is to start with the biggest number we can and work our way down. The bigger our divisor that results in a whole integer result for DOWN, is how we can calculate how many steps.

For 9, and every 8 integers moving up from 9, the formula is DOWN =  $(x-1)/8$ . These will all be exactly 3 steps, and you'll reach  $x - DOWN*2$ . so initially you'll reach 7, when you go down from 17, you'll reach 13, and when you go down from 25 you reach 19.

For 13, and every 16 integers moving up from 13, the formula is DOWN =  $(x-13)/16$ . As long as DOWN is an integer and not a float, this is 4 steps. To find out where you exit at another odd you do  $x - (DOWN*10)+8$ . If x = 13, you come out at 5, if x = 29, you come out at 11. If x = 45 you come out at 17.

For 5, we need to build a new formula. Looking at 5 we can see that it takes 5 steps and exits at 1. The correlating power of 2 is 32, that means the next integer that has exactly 5 steps is located at  $5 + 32$ . If you've filled out your spreadsheet beyond my example you can follow along. Now at 37, it also takes 5 steps and exits at 7. That's all the data we currently need.

Our divisor will be 32. You want to build this so that you either get 0, or you get 1 as a result when you perform the equation on your start number. It really depends on the difference between 5-1, and 37-7. since there is a big difference between 4 and 30, we're going to aim at making DOWN = 0 when we use 5.

DOWN =  $x-5/32$ .  $5-5 = 0$ ,  $0/32 = 0$ , DOWN = 0. Now we need to get it so we can calculate where we exit to a new odd number.  $5-4 = 1$ . so we need to make DOWN = 4, so we'll say DOWN\*n + 4. As DOWN is currently = 0, DOWN becomes 4 when we plug 5 in for x.

We then look at 30,  $30-4 = 26$ . n will be 26, so DOWN =  $(x-5)/32*26+4$ . But before we commit to this, we need to look at  $37+32$ , which is 69. 69 exits at 13.  $69-13 = 56$ . If we plug 69 in for x we get  $69-4/32 *26+4$ .  $64/32$  is  $2*26$ , is  $52+4 = 56$ . So we can calculate what DOWN is correctly, we get the amount we need to subtract from x, to get our exit integer, we know that every time we get an integer and not a float, we are dealing with every 32<sup>nd</sup> seed integer that is odd, and has 5 steps.

This pattern continues forever. The biggest formula you can get away with using and get an integer out of and not a float, THAT is the formula that gives you the result you need to count your steps and get your exit integer.

Here are some more starting formula frameworks to get you started. The following are educated guesses using the Collatz data in this tabular format, and this is what they should be, but the actual start may be different.

6 steps should be DOWN =  $x-53/64$ .

7 steps should be DOWN =  $x-21/128$ .

8 steps should be DOWN =  $x-213/256$

9 steps should be DOWN =  $x-85/512$

10 steps should be DOWN =  $x-853/1024$

11 steps should be DOWN =  $x-341/2048$

As you can see, the formulas for solving collatz are infinite. Whenever you start veering off the path from the established routes you probably need to add in more formulas. Make sure the formulas

correctly point at the entrance to the correct set, and make sure you create a formula that solves for the exit integer.

All you need to do is keep filling out your step counter, the correlating power of 2, and then start plugging in numbers in your sheet, if you fill your sheet out correctly you can just drag fill the whole thing out. If you know how to make your columns static then you'll have a much easier time figuring out the first integer where the pattern starts. As you can see, 7 steps overlaps with 21, you can find 21, which goes to 64, which then takes 7 steps to reach 1, a difference of 20. Every time you encounter an integer seed that moves you directly into the powers of 2, that can be considered a power of 4, take a note of it, this is a valid exit point for Collatz.

The only other thing we need to do is account for all the even numbers.

The formula is really simple, as all the evens go down, you want to divide your even number by the biggest power of 2 you can divide it by and get a whole integer. If you divide by 2, that's 1 step, if you divide by 32, that's 5 steps.

In conclusion, for every 8 seed integers, there are 4 odd integers, and 4 even integers. Every other odd integer starting at 3 hops up, and every other odd integer starting at 5 hops down. The amount they hop down is determined by which down formula applies where you get an integer result and not a float. Of the even integers, all of them are heading down. The number of steps they take is directly related to the biggest power of 2 that they can be evenly divided by.

This means for every 8 integers, 2 are going up, and the other 6 are traveling down. Because of the shifting nature of hops going up, and the hops coming down, your start integer will go up and down. And no matter how high it goes, it will always come back down.