International Journal of Undergraduate Research and Creative Activities

Volume 4 | Issue 1 Article 3

January 2012

Generalized Collatz Functions: Cycle Lengths and Statistics

Hayden R. Messerman
messermh@cwu.edu

Joey LeBeau
lebeauj@cwu.edu

Dominic Klyve
klyved@cwu.edu

Generalized Collatz Functions: Cycle Lengths and Statistics

Peer Review
This work has undergone a double-blind review by a minimum of two faculty members from institutions of higher learning from around the world. The faculty reviewers have expertise in disciplines closely related to those represented by this work. If possible, the work was also reviewed by undergraduates in collaboration with the faculty reviewers.

Abstract
Consider the function T (n) defined on the positive integers as follows. If n is even, T (n) = n/2. If n is odd, T (n) = 3n + 1. The Collatz Conjecture states that for any integer n, the sequence n, T (n), T (T (n)), . . . will eventually reach 1. We consider several generalizations of this function, focusing on functions which replace "3n + 1" with "3n + b" for odd b. We show that for all odd b < 400, and all integers n ≤ 106, iterating this function always results in a finite cycle of values. Furthermore, we empirically observe several interesting patterns in the lengths of these cycles for several classes of values of b.

Acknowledgements
Each of the authors was supported by the NSF-funded CWU Robert Noyce SMART Pro- gram (DUE-1035514), administered through the Center for Excellence in Science and Math- ematics Education at Central Washington University. They are grateful to Martha Kurtz for her support and for her expert administration of this grant. John Rogan, Jim Harper, and Kathy Temple read drafts of the manuscript and helped avoid some embarassing mistakes. We are grateful to two anonymous referees who carefully read our first submission and provided details and very helpful feedback. This paper is better for their efforts. The authors also wish to thank Nate Thomas, for his computational and moral support. Editor's Note: Dr. Dominic Klyve, Assistant Professor, Department of Mathematics, Central Washington University, served as faculty mentor for the authors.

1. Introduction – the $3n+1$ problem

The $3n+1$ problem concerns the following experiment: Pick any positive integer n. If n is even, divide it by 2. If n is odd, multiply it by 3 and add 1. Iterate this process by applying the same procedure to the result. Repeat this iteration process many times. The question is: what will happen in the long run?

The *Collatz Conjecture* states that no matter which integer one starts with, if one does this procedure enough times, eventually the sequence of numbers generated by it will include 1. That is, given an integer n, if we define the function $T(n)$ as

$$T(n) = \begin{cases} \frac{n}{2} & \text{if } n \equiv 0 \pmod 2 \\ 3n+1 & \text{if } n \equiv 1 \pmod 2, \end{cases}$$

[1]

and we write $T^{(k)}(n)$ to mean iterating $T(n)$ k times, then the Collatz conjecture states that for every positive integer n, there is some k such that $T^{(k)}(n) = 1$. Simply by choosing different values of n and applying $T(n)$ several times, one can be convinced that the conjecture seems to be true. Despite several decades of effort and several hundred published papers on the subject (see Lagarias, 2003, 2011), the conjecture remains unsolved[2]. Some of the best evidence in favor of the conjecture is the strong body of computational work. The current record for empirical verification of the Collatz conjecture is due to Oliveira e Silva, who has verified the conjecture for all $n < 5.76 \times 10^{18}$ (Oliveira e Silva, 2010).

Beyond the primary queston of whether iterating $T(n)$ eventually leads to 1, however, are many other questions about this function. How many iterations of $T(n)$ are needed to reach 1 for various n? How large might the intermediate values get? Does $T(n)$ behave qualitatively differently on certain classes of input (primes, odd numbers, etc.)?

Many of these questions, too, remain unsolved or only partially solved. This work is an attempt to gain new insight into these questions by generalizing $T(n)$ to a wider class of functions, and by studying their respective behavior. Consider the conjecture again. It tells us that if a number is odd, we should multiply it by 3 and add 1. Why 3? Why 1? Why should we check for divisibility by 2 and then divide? Might other numbers work? And, crucially for our work, does varying these numbers affect qualitatively the outcome of iterating these new versions of $T(n)$?

2. Generalized functions

The idea of generalizing the function $T(n)$ dates back at least to Hasse (1975), who suggested a generalization largely similar to the one below,

[1]The notation $n \equiv 0 \pmod 2$ means that n divided by 2 will have a remainder of 0; more generally $a \equiv b \pmod m$ means that a divided by m will have remainder b.

[2]Gerhard Opfer has submitted a paper containing a proported proof, and the mathematics community is currently waiting to see whether the proof will stand under close scrutiny

https://digitalcommons.cwu.edu/ijurca/vol4/iss1/3
DOI: 10.7710/2168-0620.1002

DOI: http://dx.doi.org/10.7710/2168-0620.1002

and then gave some probabilistic arguments about the behavior of these generalized functions. In particular, Hasse suggested that for a fixed $m > d \geq 2$, one could define the map

$$D_d(x) := \frac{mx + f(r)}{d} \text{ if } x \equiv r \pmod{d},$$

where $f(r) \equiv -mr \pmod{d}$. In our generalization, we change only the constant being added in the fraction's numerator, in order to allow ourselves more freedom in varying the original question. In particular, we define $T(n;m,a,b)$ as

$$T(n;m,a,b) = \begin{cases} \frac{n}{m} & \text{if } n \equiv 0 \pmod{m}; \\ an + b & \text{if } n \not\equiv 0 \pmod{m}. \end{cases}$$

Some experimentation quickly reveals that some values of m, a, and b give rise to uninteresting functions. For example, $T(n;2,3,2)$ triples any odd number and adds 2 – giving another odd number. This continues until the iterates become arbitrarily large; see Theorem 2.1. Other functions may have some starting values for which they go to infinity, but others for which they converge. Still others might send all numbers into one of many possible "cycles".

We note also that many other generalizations of this function are possible, but we believe the our version, based on that of Hasse, is the most general which "preserves the spirit" of the original problem. This generalization of Hasse has been studied by Allouche (1979), Heppner (1978), Möller (1978), Metzger (1999), and Lagarias (1990).

2.1. **Cycles and divergent sequences.** The original Collatz function seems (empirically) to continue iterating an input until the values reach the cycle $4 \to 2 \to 1 \to 4$. One of the first interesting things we found was that some values of m, a, and b, define functions which have more than one cycle. Additionally, some of these cycles are quite long. For example, consider the effect of modifying the original $T(n)$ only slightly, by changing the rule if n is odd to "multiply by 3 and add 5". We represent this function as $T(n;2,3,5)$. Under this function, the integers $19 \to 62 \to 31 \to 98 \to 49 \to 152 \to 76 \to 38 \to 19$ form a *cycle*. In fact, this function has six distinct cycles which include at least one integer less than 10^5.

Other functions take inputs which do not converge at all. For example, consider the function $T(n;2,5,3)$, with starting value $n = 5$. The pattern of iterates begins $5 \to 28 \to 14 \to 7 \to 38 \to 19 \to 98 \to 49 \to 248 \to 124 \to 62 \to 31 \to 138 \to 79 \to 398 \to 199 \to 998 \to 499 \to \cdots$. The pattern in these integers seems fairly random, as the values move up and down. The general trend, however, is that the sequence is growing larger – the 1000th iterate is 6.16×10^{41}. Such inputs are said to be *divergent*.

2.2. **Previous Work.** The class of functions $T(n;2,m,b)$ was studied by Crandall (1978), who proved several interesting results. In particular, he conjectured that aside from $(m,b) = (3,1)$, every function $T(n;2,m,b)$

DOI: http://dx.doi.org/10.7710/2168-0620.1002

has at least one cycle that does not reach 1; he proved this for all $b \geq 3$, and for the pairs $(m,b) \in \{(5,1),(181,1),(1093,1)\}$. Additionally, function of the form $T(n;2,3,b)$ (the $3x+b$ functions) have been studied by Belaga and Mingotte in (Belaga & Mignotte, 1998) – see also their unpublished reports in (Belaga & Mignotte, 2000) and (Belaga & Mignotte, 2006).

2.3. **An interesting class of functions.** After some initial experimentation with various values of a, b, and m, we noticed that while the behavior of the functions varied widely, the functions with $a = 3$ and $m = 2$ behaved in a somewhat uniform manner. Specifically, if we fixed those values and let b vary through odd values, we could find no starting value n which did not end up in some cycle (see Theorem 4.1). However, when b was even, we saw very different behavior. In fact, we have the following:

Theorem 2.1. *Let b be a positive even integer and n a postiive odd integer. Then*

$$\lim_{k\to\infty} T^{(k)}(n;2,3,b) = \infty.$$

Proof. Since n is odd, set $n = 2m+1$. Then

$$\begin{aligned} T(n;2,3,b) &= 3n+b \\ &= 3(2m+1)+b \\ &= 6m+b+1, \end{aligned}$$

which is odd (since b is even). Futhermore, since $6m+b+1 > 2m+1 = n$, we will always have that $T(n;2,3,b) > n$. □

For this reason we shall restrict our attention to functions $T(n;2,3,b)$ with odd b. Beyond this observation, though, no clear pattern was present. Some of these functions (like the standard function $T(n)$) seemed to send all integers to the same cycle. Others had many different cycles which appeared. It was to a deeper understanding of the structure and behavior of cycles that we next turned our efforts.

3. Cycle Statistics

3.1. **Some definitions.** Because in the rest of this work, we shall be studying functions of a particular type, it may be useful to assign new notation to these. Let $T(n;2,3,b)$ be denoted by

$$C_b(n) = \begin{cases} \frac{n}{2} & \text{if } n \equiv 0 \pmod 2; \\ 3n+b & \text{if } n \not\equiv 0 \pmod 2. \end{cases}$$

We shall refer to these $C_b(n)$ collectively as *Collatz Functions*. The Collatz function $C_1(n)$, then, corresponds to the original function described in the introduction. Recall also that we restrict our attention only to the case where b is odd.

In order to describe properties of these Collatz functions, we shall need a few definitions.

Definition 3.1. *The* **cycle number** *of a Collatz function C_b is the number of distinct cycles created by iterating C_b.*

We note first the crucial fact that for no function C_b is the cycle number known. One way to state the original Collatz conjecture is to say that the

https://digitalcommons.cwu.edu/ijurca/vol4/iss1/3
DOI: 10.7710/2168-0620.1002

DOI: http://dx.doi.org/10.7710/2168-0620.1002

cycle number of C_1 is 1, but of course this has not been proved.

In the following, when we refer to the cycle number, we always mean the number of distinct cycles we have found empirically: for $b < 100$, this is the number of cycles observed by testing all inputs up to 10^6; for $100 < b < 400$, we tested inputs to up 10^5. The cycle number for some of the C_b are given in the following table:

b	Cycle number of C_b
1	1
3	1
5	6
7	2
9	1
11	3
13	10

Table 3.1: Data on cycle numbers of C_b for small b

Note that like C_1, C_3 also seems to have a cycle number of 1. The function C_5, however, behaves very differently. It has cycle number 6. There is once again a cycle containing 1, namely $1 \to 8 \to 4 \to 2 \to 1$. Using 3 as our starting value, however, gives $3 \to 14 \to 7 \to 26 \to 13 \to 40 \to 20 \to 10 \to 5 \to 20$ – these last three numbers represent a cycle distinct from the first. Others are much longer. One cycle of C_5 begins with 187, then passes through 43 other integers before returning to 187. These differences suggest that it will be useful to have the following:

Definition 3.2. *The* **cycle length** *of any cycle is the number of distinct integers contained within the cycle.*

Given that some cycles are quite long, we shall also need a way to represent them without writing out every integer in the cycle. For this, we shall use the following:

Definition 3.3. *For a given Collatz function C_b and a given cycle, the* **cycle minimum** *is the least integer in the cycle.*

Since no integer can be in more than one cycle of a given C_b, a cycle minimum uniquely determines a cycle of a given C_b. For this reason we shall often use the value of the cycle minimum to describe the cycle itself.

Definition 3.4. *For a given Collatz function C_b and a given cycle, the* **cycle gravity** *is the proportion of integers which end up in the cycle (if such a proportion exists).*

The motivation for this choice of terminology is simple: loosely, the greater the gravity, the more integers are "pulled in" to the cycle. We should note, however, that it is not clear that this is well-defined. Let $P_{b,r}(n)$ be the number of integers not greater than n which end up in a cycle with cycle minimum r after iteration by C_b. The cycle gravity is $\lim_{n\to\infty} P_{b,r}(n)/n$, if the limit exists.

DOI: http://dx.doi.org/10.7710/2168-0620.1002

4. Computations and Observations

4.1. **General experiments.** For every value of $b \leq 99$, we tested all inputs $n \leq 10^6$. For each of these, we first found all cycles which are entered by any of the n. The first important discovery is that every value of n does indeed enter a cycle. In fact, we verified the following generalization of the Collatz Conjecture.

Theorem 4.1. *For all odd $b \leq 99$ and all $n \leq 10^6$, the iterates of $C_b(n)$ eventually converge to some cycle.*

For each cycle, we recorded several pieces of data. We first recorded the cycle minimum and the cycle length. Metzger (1999) observed that it is often useful to measure a cycle not by its total number of entries, but by the number of odd entries, which correspond to its number of multiplications. We also computed this value, but we found it not to be more useful than our definition, and we do not report that value here (but see the accompanying tables online for complete data). Finally, we recorded each cycle's gravity, in this case estimated using the proportion of integers $n \leq 10^6$ which enter that cycle. Part of this data (for $b \leq 13$) is given in Table 4.1.

We attempted to find some relationship between a cycle's length and its gravity. No pattern is apparent, though we encourage readers to look for one. (All data for $b < 400$ are available in the accompanying tables online.)

b	**Cycle min**	**Length**	**Gravity**
1	1	3	1.00000
3	3	3	1.00000
5	1	4	0.14099
	5	3	0.20000
	19	8	0.49642
	23	8	0.09297
	187	44	0.03253
	347	44	0.03709
7	5	6	0.85714
	7	3	0.14285
9	9	3	1.00000
11	1	8	0.19752
	11	3	0.09091
	13	22	0.71157
13	1	5	0.47686
	13	3	0.07692
	131	39	0.25215
	211	13	0.03340
	227	13	0.01911
	251	13	0.01971
	259	13	0.03885
	283	13	0.02541
	287	13	0.04301
	319	13	0.01459

Table 4.1: Data on cycles for small b

4.2. **Cycles C_b where b is a multiple of 5.** Despite not seeing any clear patterns between cycle gravity and the other statistics, we found several curious relationships between cycle lengths and the values of b, which we believe to be true for all b. Consider Table 3, which gives information similar to that in Table 2, but is restricted only to those b which are (odd) multiples of 5.

Particularly curious is the fact that each of these values of b has exactly two cycles of length 44. In fact, this pattern continued as far as we were able to test. Based on the limits on our experimentation, this gave us the following:

Theorem 4.2. *For all integers* $b < 400$ *which are odd multiples of* 5*,* C_b *has at least two cycles with cycle length* 44 *and two with cycle length* 8*.*

(See Conjecture 5.3 below for more details).

b	**Cycle min**	**Length**	**Gravity**
5	1	4	0.14099
	5	3	0.20000
	19	8	0.49642
	23	8	0.09297
	187	44	0.03253
	347	44	0.03709
15	3	4	0.2
	3	4	0.141344
	15	3	0.066667
	57	8	0.495195
	69	8	0.093219
	561	44	0.032591
	1041	44	0.037651
25	5	4	0.028286
	25	3	0.04
	95	8	0.099078
	115	8	0.018673
	17	12	0.376009
	7	24	0.423991
	935	44	0.006457
	1735	44	0.007506

Table 4.2: Cycles for which b is divisible by 5

4.3. **Cycles C_b for b divisible by other primes.** Studies of values of b divisible by some other primes generated other interesting results. We give some of these in the following two theorems.

Theorem 4.3. *For all integers* $b < 400$ *which are odd multiples of* 13*,* C_b *has seven cycles with cycle length* 13*.*

Theorem 4.4. *For all integers* $b < 400$ *which are odd multiples of* 29*,* C_b *has one cycle with cycle length* 106*.*

These observations seem to point experimentally to a more general property; namely that if $n > 1$ is an integer, and C_b has a cycle of length k, then C_{nb} will also have a cycle of length k. In fact, this is the case, as we shall prove in the next section.

5. Results

Before we state our results, it will be useful to have the following definition:

Definition 5.1. *A cycle* $\{x_1, x_2, \ldots x_k\}$ *of a function* C_b *is said to be* primitive *if the greatest common divisor of all the* x_i *is* 1*.*

Using this, it is easy to prove the following:

Theorem 5.2. *Let* m *be the cycle minimum of a primitive cycle of length* k *under the function* C_b*. Then for any integer* $n > 1$*,* nm *will be the cycle minimum of a (imprimitive) cycle of length* k *under the function* C_{nb}*.*

Proof. The idea of the proof is that the function C_{nb} preserves the primitive cycle in C_b, and merely scales it up. For $C_{nb}(nm) = 3nm + nb =$

DOI: http://dx.doi.org/10.7710/2168-0620.1002

$n(3m+b) = nC_b(m)$. Therefore after futher interation, we see that

$$C_{nb}^k(nm) = nC_b^k(m) = m,$$

since m was the minimum of a cycle of length k. □

From this theorem, we can deduce three corollaries which parallel the three theorems above:

Corollary 5.3. *If b is an odd multiple of 5, C_b will have at least two cycles with cycle length 44 and two cycles with cycle length 8.*

In fact, we observe more. Usually these two cycles have cycle minima greater than any other cycle given by C_b. We found this to be true for all b except those which are multiples of 29. we cannot prove, nor are we convinced, that C_{5b} will always have exactly two cycles of length 44.

Corollary 5.4. *If b is an odd multiple of 13, C_b will have at least seven cycles of length 13.*

Corollary 5.5. *If b is an odd multiple of 29, C_b will have at least two cycles of length 106.*

Finally, we note that of all $b < 400$, the only C_b with cycle number 1 are $b = 1, 3, 9, 27, 81, 243$. From this, we conjecture the following:

Conjecture 5.6. *The function C_b will have precisely one cycle if and only if b is a power of 3.*

We note that it is easy to see that such functions always have at least one cycle; that the function C_{3^n} has a cycle with cycle minimum 3^n follows from Theorem 5.2.

6. Conclusions

In one sense, the generalization of the Collatz function to the functions C_b leads to very similar behavior to that seen in the original function – namely, iterating any value leads to a cycle (rather than diverging). This suggests at least that the original function is not "special", but rather that its behavior follows from more general principles. However, the fact that the number of cycles differs as b changes does give evidence that there are underlying irregularities behind the more systematic behavior. This, together with the striking patterns in behavior for values of b which are multiples of 5, 13, and 29, leads us to believe that there is considerable benefit to studying these generalized functions.

7. Note

We would like to acknowledge one of our reviewers who pointed out that many of our results are not as novel as we thought, in particular, for directing our attention to prior work of Lagarias (1990) that our work extends. We were unaware of this work at the time of our own studies. To our knowledge, our data on cycle gravity is still new.

8. References

Allouche, J. (1979). Sur la conjecture de "Syracuse-Kakutani-Collatz". In *Seminar de théorie des nombres. exp* (pp. 1978–1979).

https://digitalcommons.cwu.edu/ijurca/vol4/iss1/3
DOI: 10.7710/2168-0620.1002
DOI: http://dx.doi.org/10.7710/2168-0620.1002

Belaga, E., & Mignotte, M. (1998). Embedding the 3x+ 1 conjecture in a 3x+ d context. *Experimental Mathematics*, *7*, 145–152.

Belaga, E., & Mignotte, M. (2000). Cyclic structure of dynamical systems associated with $3x+d$ extensions of collatz problem. Research report, available at http://hal.archives-ouvertes.fr/hal-00129656/.

Belaga, E., & Mignotte, M. (2006). The collatz problem and its generalizations: Experimental data. table 1. primitive cycles of $(3n+d)-$ mappings. Institut de Recherche Mathématique Avancée de Strasbourg, Strasbourg.

Crandall, R. (1978). On the" 3x+1" problem. *Mathematics of Computation*, 1281–1292.

Hasse, H. (1975). Unsolved problems in elementary number theory. *Lectures at U Maine (Orono), Spring*, 23–33.

Heppner, E. (1978). Eine Bemerkung zum Hasse-Syracuse-Algorithmus. *Archiv der Mathematik*, *31*(1), 317–320.

Lagarias, J. (1990). The set of rational cycles for the 3x+ 1 problem. *Acta Arith*, *56*, 33–53.

Lagarias, J. (2003). The 3x+ 1 problem: An annotated bibliography (1963–1999)(sorted by author). *Arxiv preprint math/0309224*.

Lagarias, J. (2011). The 3x+ 1 problem: An annotated bibliography, ii (2000–). *Arxiv preprint math.NT/0608208*.

Metzger, K. (1999). Beitrage-Zyklenbestimmung beim (bn+ 1)-Algorithmus. *Praxis der Mathematik*, *41*(1), 25–27.

Möller, H. (1978). Uber Hasses Verallgemeinerung des Syracuse-algorithmus (Kakutani's problem). *Acta Arith*, *34*, 219–226.

Oliveira e Silva, T. (2010). Empirical verification of the $3x+1$ and related conjectures. In J. C. Lagarias (Ed.), *The ultimate challenge: The $3x+1$ problem* (pp. 189–207). Providence, Rhode Island, USA : American Mathematical Society.

DOI: http://dx.doi.org/10.7710/2168-0620.1002