The Cyclic Nature (and Other Intriguing Properties) of Descriptive Numbers

Trevor Scheopner
University of Kansas

Abstract

In this paper we discuss the Descriptive Numbers which are in definition almost identical to J. Conway's popular Look and Say numbers and analyze the significant difference in the behavior of the two sequences. Whereas the terms in Conway's sequence grow in length indefinitely, the Descriptive Numbers do not exhibit the same property. Instead, the length of Descriptive Sequence terms is bounded. The result of trying to relate the lengths of successive terms in a sequence of Descriptive Numbers we show is proving that all Descriptive Sequences are cyclic regardless of numerical base or initial condition.

1 Introduction

A sequence of numbers which has been explored before in recreational mathematics, called the "Descriptive Numbers" by the Online-Encyclopedia of Integer Sequences [1], begins 1, 11, 21, 1112, 3112, and so on. The descriptive numbers continue by the process wherein the next term is the number of occurrences of the least digit in the previous term followed by that digit, followed by the number of occurrences of the next least digit and then by that digit, and so on for each unique digit in the previous term. So, because there is 1 "one" in "1", the next term in the sequence is 11. Then, because there are 2 "ones" in "11", the next term in the sequence is 21. The digits are ordered increasingly; so, the next terms after that are 1112 (1 "one" and 1 "two") and 3112 (3 "ones" and 1 "two"). This sequence is particularly interesting because regardless of what number is chosen as the starting point of the sequence, the sequence eventually forms a complete loop and repeats a few values cyclically thereafter.

For example, the initially given sequence progresses further as 211213, 312213, 212223, 114213, 31121314, 41122314, 31221324, 21322314. The last of these numbers, 21322314, maps to itself. So, that value is repeated for every future term of the sequence. Although not every initial condition reaches a number which maps to itself, every initial condition does reach a cycle of some length. The case of a number mapping to itself is a cycle of length 1. Not only does every initial condition in base 10 lead to a cycle but every initial condition in any numerical base leads to a cycle. The proof that the descriptive numbers are cyclic regardless of base and of initial condition is the principal result of this paper.

These numbers in base 10 specifically have been explored before by V. Bonstein and A. Fraenkel [2] using different methods. A more popular and related sequence called the "Look and Say Numbers" was explored by J. Conway in a famous paper of his, "The Weird and Wonderful Chemistry of Audioactive Decay" [3]. The sole structural difference between the Descriptive Numbers in base 10 and the Look and Say Numbers is that the former describes the total number of each digit present in the previous number, ordered as those digits increase while the Look and Say Numbers iterate from the beginning of the sequence and specify how many of each digit occur sequentially, followed by that digit. For example, the Look and Say sequence with initial condition 1 is 1, 11 (1 "one"), 21 (2 "ones"), 1211 (1 "two" followed by 1 "one"), 111221 (1 "one" followed by 1 "two" followed by 2 "ones"), 312211, 13112221, and so on. As this sequence continues it grows indefinitely. The relationship between the Descriptive Numbers and the Look and Say Numbers is discussed further in Section 4.4.

2 Definitions

Definition 2.1: Base b String

1. Let a base b string x be a sequence x_0, x_1, ..., x_n of finitely many natural numbers represented in base b such that for all x_m, where $m \in \mathbb{N}$ such that $0 \leq m \leq n$, $0 \leq x_m \leq b$. Let a base b string x be represented as "$x_0x_1x_2...x_n$". Examples: "1", "11", "21", "0099", "05", "", and "15432598765101" are all base 10 strings. "10" and "1001" are base 2 strings.

2. Let the length of a base b string x be the least $m \in \mathbb{N}$ such that x_m does not exist. Let the length of x be denoted by $l(x)$. Examples: $l(\text{"1"}) = 1$, $l(\text{"0099"}) = 4$, $l(\text{""}) = 0$, and $l(\text{"15432598765101"}) = 14$.

3. Let two base b strings x and y be considered equal if $l(x) = l(y)$ and $x_m = y_m$ for all $m \in \mathbb{N}$ such that $0 \leq m < l(x)$. For example, "11120099" = "11120099".

4. Let the concatenation of two base b strings x and y equal z and be denoted $x \oplus y = z$ if (i) $l(z) = l(x) + l(y)$, (ii) $z_m = x_m$ for all $m \in \mathbb{N}$ such that $0 \leq m < l(x)$, and (iii) $z_m = y_{m-l(x)}$ for all $m \in \mathbb{N}$ such that $l(x) \leq m < l(z)$. Also let the iterative concatenation of many base b strings be denoted by $\bigoplus$ (similarly to sigma summation). Examples:

$$\text{"1112"} \oplus \text{"0099"} = \text{"11120099"}$$
$$\text{If } f(n) = \text{"}n\text{", then } \bigoplus_{n=0}^{9} f(n) = \text{"0123456789" in base 10}$$

5. Let the content of a base b string x be the set of all values assumed in the sequence of x and be denoted $Con(x)$. Examples: $Con(\text{"1"}) = Con(\text{"11"}) = \{1\}$, $Con(\text{"114213"}) = \{1, 2, 3, 4\}$, and $Con(\text{"15432598765101"}) = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

6. Let the multiplicity of an element e of the content of a base b string x be the number of times that e occurs in the sequence x_0, x_1, ..., x_n and be denoted $Mult(x, e)$. Examples: $Mult(\text{"1"}, 1) = 1$, $Mult(\text{"11"}, 1) = 2$, $Mult(\text{"11"}, 2) = 0$, and $Mult(\text{"15432598765101"}, 5) = 3$.

7. Let $T_b(x)$ be called the description of the base b string x and defined by

$$T_b(x) := \bigoplus_{k \in Con(x)} (Mult(x, k) \oplus k).$$

Examples:

$$T_{10}(\text{"1"}) = \text{"11"} \qquad T_{10}(\text{"11"}) = \text{"21"} \qquad T_{10}(\text{"21"}) = \text{"1112"}$$
$$T_{10}(\text{"1112"}) = \text{"3112"} \qquad T_{10}(\text{"1111111111"}) = \text{"101"} \qquad T_2(\text{"1001"}) = \text{"100101"}$$

Definition 2.2: Descriptive Base b Sequence

1. Let a descriptive base b sequence $d_0, d_1, ...$ be a sequence of base b strings such that $d_{n+1} = T_b(d_n)$ for all $n \in \mathbb{N}$. Examples: "1", "11", "21", "1112", "3112", ... and "05", "1015", "102115", "10311215", ... are both descriptive base 10 sequences.

2. Let the term d_0 of a descriptive base b sequence d be called the initial condition of d. Examples: The initial condition of the first sequence given in the previous definition is "1". The initial condition of the second sequence given in the previous definition is "05".

3. Let a descriptive base b sequence d be called cyclic if there exists $m, n \in \mathbb{N}$ such that $d_m = d_n$ and $m \neq n$. Example: Let d be the descriptive base 10 sequence with initial condition "1". Then, d is cyclic because $d_{13} = d_{14}$.

3 Proof of Principal Result

Theorem 3.1: All descriptive sequences are cyclic.

First, an inequality relating the lengths of successive terms in a descriptive sequence will be found. Then, this inequality will be used to show that the length of any term in the descriptive sequence is bounded. Then, the boundedness of the length will be used to deduce that any descriptive sequence is cyclic. Each of these statements taken as a lemma makes the theorem a simply their combination.

Lemma 3.1.1: Let d_n be a descriptive base b sequence and let L_n be a sequence such that $L_n = l(d_n)$ for all n. Then, $L_{n+1} \leq b\log_b(L_n) + 2b$.

Proof: The number of digits which the multiplicity of k in d_n contributes to the length of d_{n+1} is the number of digits of $Mult(d_n, k)$ plus the number of digits of k. The number of digits of k is always 1. The number of digits of $Mult(d_n, k)$ is $\lfloor \log_b(Mult(d_n, k)) \rfloor + 1$. (Thoughout this paper, $\lfloor x \rfloor$ will be used to denote the greatest integer less than or equal to x.) So,

$$L_{n+1} = l(d_{n+1}) = \sum_{k \in Con(d_n)} (\lfloor \log_b(Mult(d_n, k)) \rfloor + 2).$$

$Mult(d_n, k) \leq l(d_n)$ for all k and $\log_b(x)$ is increasing for all x. So,

$$L_{n+1} = \sum_{k \in Con(d_n)} (\lfloor \log_b(Mult(d_n, k)) \rfloor + 2) \leq \sum_{k \in Con(d_n)} (\lfloor \log_b(l(d_n)) \rfloor + 2).$$

$\lfloor x \rfloor \leq x$ for all x. So,

$$L_{n+1} \leq \sum_{k \in Con(d_n)} (\lfloor \log_b(l(d_n)) \rfloor + 2) \leq \sum_{k \in Con(d_n)} (\log_b(l(d_n)) + 2).$$

The new expression inside the sum is independent of k so it simplifies to

$$L_{n+1} \leq |Con(d_n)|(\log_b(L_n) + 2).$$

$|Con(d_n)| \leq b$. So,

$$L_{n+1} \leq b\log_b(L_n) + 2b.$$

Call this inequality the string length inequality.

Lemma 3.1.2: Given the string length inequality, the sequence L_n is bounded.

Proof: Let $y_b(x) = b\log_b x + 2b$, $F_{n+1} = y_b(F_n)$, and $F_0 = L_0$. Because $y_b(x)$ is increasing, if $L_k \leq F_k$ then $y_b(L_k) \leq y_b(F_k)$. Then, because $y_b(F_k) = F_{k+1}$ (by the definition of F_n) and $L_{k+1} \leq y_b(L_k)$ (by the string length inequality), we have that if $L_k \leq F_k$ then $L_{k+1} \leq F_{k+1}$. And thus, by induction, because $L_0 = F_0$, $L_n \leq F_n$ for all n. Now, if F_n is bounded above, then L_n is bounded above. L_n is bounded below by 0. Hence, if F_n is bounded above, L_n is bounded.

Notice that F is a simple recurrence relation. By the definition of F, a fixed point for F must satisfy $F_n = y_b(F_n)$. Consider $g(b, x) = y_b(x) - x = b \log_b x + 2b - x$. Then, the ordered pairs (b, x) which satisfy $g(b, x) = 0$ are the fixed points for F in base b. By the Implicit Function Theorem, x is parametrized by b (equivalently, for that base b there is only one fixed point and it is located at $x(b)$) whenever $\partial g/\partial x \neq 0$.

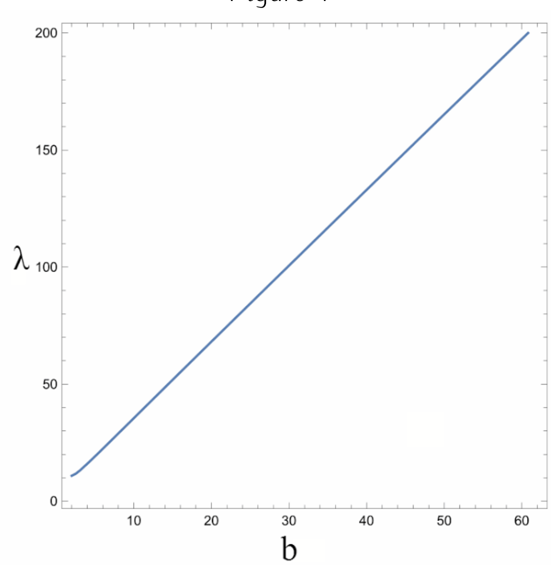
$$\frac{\partial g}{\partial x}(b, x) = \frac{b}{x \ln b} - 1$$

Hence, $\partial g/\partial x = 0$ if and only if $x = b/\ln b$. Consider now the order pairs $(b, b/\ln b)$ which satisfy $g(b, x) = 0$, recalling that $b > 1$ because b is a numerical base.

$$\begin{aligned} g\left(b, \frac{b}{\ln b}\right) =& \frac{b}{\ln b} \ln\left(\frac{b}{\ln b}\right) + 2b - \frac{b}{\ln b} = 0 \\ & \frac{b}{\ln b}(3\ln b - \ln\ln b - 1) = 0 \\ & \ln\frac{b^3}{\ln b} = 1 \\ & b^3 - e\ln b = 0 \end{aligned}$$

The left side of this equation is 1 when b is 1 and its derivative with respect to b is $3b^2 - e/b$ which for $b > 1$ is always positive. Hence, the left side of this equation never equals 0 and there are no b for which x is not parametrized by b. Let the x solution to $g(b, x) = 0$ for a particular b be denoted λ_b. The graph of λ_b as it is implicitly defined by b is shown below in Figure 1. The graph looks deceptively linear.

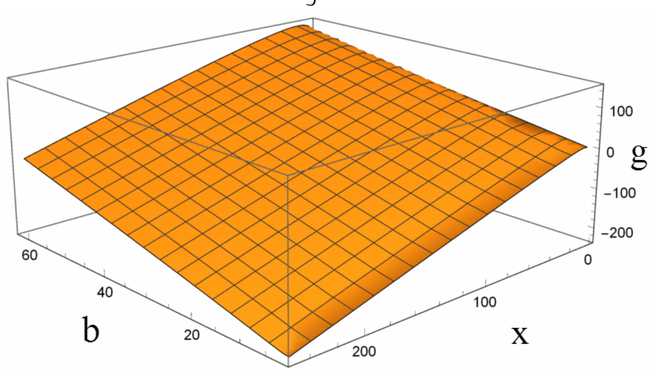
Figure 1



Notice from the definition of $g(b, x)$ or below in Figure 2 that for arbitrarly large x, $g(b, x)$ becomes an arbitrarly large negative number. Notice also that, as demonstrated earlier by the Implicit Function

Theorem, the stationary point with respect to x of $g(b, x)$ is never also the unique x which satisfies $g(b, x) = 0$ (λ_b). Hence, $g(b, x) > 0$ for $x < \lambda_b$, $g(b, x) = 0$ for $x = \lambda_b$, and $g(b, x) < 0$ for $x > \lambda_b$. Consequently, because $g(b, x) = y_b(x) - x$, $y_b(x) > x$ for $x < \lambda_b$, $y_b(x) = x$ for $x = \lambda_b$, and $y_b(x) < x$ for $x > \lambda_b$.

Figure 2



Consider the case of $F_0 \leq \lambda_b$. Because $y_b(x)$ is increasing and λ_b is a fixed point for $y_b(x)$, if $F_k \leq \lambda_b$ for some k then, $F_{k+1} = y_b(F_k) \leq y_b(\lambda_b) = \lambda_b$. Hence, if $F_k \leq \lambda_b$ then $F_{k+1} \leq \lambda_b$. Thus, by induction, if $F_0 \leq \lambda_b$, F_n is bounded above by λ_b.

Now consider the case of $F_0 > \lambda_b$. Because $y_b(x)$ is increasing and λ_b is a fixed point for $y_b(x)$, if $F_k > \lambda_b$ for some k then, $F_{k+1} = y_b(F_k) > y_b(\lambda_b) = \lambda_b$. Hence, if $F_k > \lambda_b$ then $F_{k+1} > \lambda_b$. Thus, by induction, if $F_0 > \lambda_b$, F_n is bounded below by λ_b. Now, because $y_b(x) < x$ whenever $x > \lambda_b$, $y(F_k) = F_{k+1} < F_k$. Hence, if $F_0 > \lambda_b$, F_n is strictly decreasing and thus bounded above by F_0.

Hence, regardless of the value of F_0, F_n is bounded above. Thus, L_n is bounded.

Lemma 3.1.3: The boundedness of L_n is sufficient to conclude that d_n is cyclic.

Proof: Let $\beta_b = \lfloor \lambda_b \rfloor + 1 > \lambda_b$ if $L_0 < \lambda_b$ and let $\beta_b = L_0$ if $L_0 \geq \lambda_b$. Then, β_b is an upper bound for L_n. Note that in a base b string with length equal to β_b each digit has b possible independent values hence there are b^{β_b} possible such base b strings. Consequently there are

$$1 + b + b^2 + ... + b^{\beta_b} = \frac{b^{\beta_b+1} - 1}{b - 1} < b^{\beta_b+1}$$

different base b strings with length less than or equal β_b. So, consider $d_{b^{\beta_b+1}}$. If any pair of values of d_n preceeding that value in the sequence are equal, then d_n is cyclic. If not, then all of the preceeding values are unique which, because there are fewer possible unique values for d_n than b^{β_b+1}, means that the value of $d_{b^{\beta_b+1}}$ must have occurred previously in the sequence. Hence, d_n is cyclic. **Q.E.D.**

4 Observations and Room for Further Investigation

4.1 Cycle Length

After some initial exploration of descriptive numbers one might conclude that all sequences reach a string which maps to itself. However, this is not the case. When the first 1000000 initial conditions are checked in base 10 by a computer (which iterated through sequences with initial conditions as “0", “1", ... “9", “00", ... “09", “11", ... “19", ..., “88", “89", “99", “000", ... as there is no use having any non-increasing digit sequence as an initial condition due to the order independence of the description process), it is revealed that the least initial condition with a cycle of length greater than 1 is “04". The cycle length of “04" is 2. The least initial condition with a cycle of length 3 is “05". The descriptive sequences for both are shown below.

“04” ↦ “1014” ↦ “102114” ↦“10311214” ↦ “1041121314” ↦ “1051121324” ↦ “104122131415” ↦ “105122132415” ↦ “104132131425” ↦ (“104122232415” ↦ “103142132415”) (Last 2 elements form a cycle)

“05” ↦ “1015” ↦ “102115” ↦ “10311215” ↦ “1041121315” ↦ “105112131415” ↦
“106112131425” ↦ “10512213141516” ↦ “10612213142516” ↦ “10513213141526” ↦
(“10512223142516” ↦ “10414213142516” ↦ “10512213341516”) (Last 3 elements form a cycle)

The first cycle of length 3 is only the 16$^{\text{th}}$ decimal string checked. Strangely though, no cycles of length greater than 3 occur in the first 1000000 decimal strings. Also, 3-cycles are not rare at all. Of the first million decimal strings, 140806 of them result in a 3-cycle.

The maximum cycle length in any base is bounded as a consequence of the fact that the length of a sequence is bounded. However, the distribution of each cycle length as a function of the base is currently unexplored. My analysis has so far failed to even produce a 4-cycle or greater in base 10 or a 2-cycle or greater in base 2.

4.2 Pre-Cycle Length

Another characteristic of intrigue is how long it takes for a descriptive sequence to complete its first cycle (or equivalently, how many unique numbers are in the sequence). Similarly to cycle length, the extreme initial condition (in this case, the initial condition which takes the longest to repeat) in the first million base 10 strings occurs early on. That initial condition is “0099" and its descriptive sequence contains 20 unique strings. Its descriptive sequence is given below.

“0099” ↦ “2029” ↦ “102219” ↦ “10212219” ↦ “10313219” ↦ “1031122319” ↦ “1041222319” ↦
“103132131419” ↦ “105112331419” ↦ “10511223141519” ↦ “10612213142519” ↦
“1051321314151619” ↦ “1071122314251619” ↦ “106132131415161719” ↦ “108112231415261719” ↦
“10713213141516171819” ↦ “10911223141516271819” ↦ “10813213141516171829” ↦
(“10812223141516172819” ↦ “10713213141516172819”) (Last 2 elements form a cycle)

In the first million base 10 strings, 7 have pre-cycle length of 20. Similarly to maximum cycle length, the maximum pre-cycle length in a given base is bounded. The distribution of pre-cycle lengths and whether it varies across the initial condition in a simple predictable manner is currently unknown.

4.3 Funneling and Isolation

Consider the descriptive sequences with initial conditions “0" and “1" in binary, given below.

“0” ↦ “10” ↦ “1011” ↦ “10111” ↦ “101001” ↦ “110111” ↦ “101011” ↦ “1001001” ↦ “1000111” ↦ (“1101001”) (Last element forms of cycle)

"1" ↦ "11" ↦ "101" ↦ "10101" ↦ "100111" ↦ "1001001" ↦ "1000111" ↦ ("1101001") (Last element forms a cycle)

Interstingly, these two initial conditions lead to the same cyclic element. In binary, all of the first 1000000 initial conditions (all of the initial conditions which have been checked) eventually lead to this exact final element, '1101001", except for a single one. This funneling phenomenon occurs in all of the bases I have investigated numerically but is most significant in binary. Binary is also interesting because of that sole exception which is one of the only two isolated nodes found thusfar. Consider "111". It maps to itself. "1101001" also maps to itself. However, many different elements map to "1101001" but it is easily checked that only "111" maps to "111". Hence, "111" is completely isolated. The other isolated node is even more interesting as it is isolated in every base greater than 2. That node is "22". The existence of other isolated nodes and the extent of funneling in other bases is currently an open question.

4.4 Relation to the Look and Say Numbers

Regardless of the base, once a certain digit is in a sequence it is in all future terms of the sequence. Consider the descriptive sequence with the initial condition "09" in base 10, given below.

"09" ↦ "1019" ↦ "102119" ↦ "10311219" ↦ "1041121319" ↦
"105112131419" ↦ "10611213141519" ↦ "1071121314151619" ↦
"108112131415161719" ↦ "10911213141516171819" ↦ "10101121314151617 1829" ↦
"20101221314151617 1819" ↦ "20913213141516171819" ↦ "10812223141516171829" ↦
("10714213141516172819" ↦ "10812213241516271819") (Last two elements form a cycle)

Notice that this initial condition slowly generates every digit in base 10 by overflowing with ones. Incorporating 0 in an initial condition produces this effect. It causes an additional 1 to appear in every term of the sequence producing an increasing number of ones rather than a nearly constant number of ones across terms until the different digits available in the base are exhausted. This type of behavior is somewhat similar to the behavior of the Look and Say Numbers in that the Look and Say Numbers have "atoms" which are self-reproducing and generate structure in their vicinity in the sequence while this sequence essentially has the initial condition as the atom and the structure it produces is the complete list of digits available in the base. However, this also illustrates the fundamental difference between the Descriptive Numbers and the Look and Say Numbers. The Look and Say Numbers have atoms which generate structure in their vicinity because it is order dependent. The Descriptive Numbers are order independent and hence a "mixing" occurs. If x and y are strings, then the Look and Say algorithm applied to x concatenate y equals the Look and Say algorithm applied to x concatenated with the Look and Say algorithm applied to y whenever the last digit of x is different from the first digit of y and the two only differ at the point of concatenation between x and y when those two digits are the same. However, the Descriptive algorithm applied to x concatenate y generally bears no resemblence to their individual Descriptions due to the mixing which occurs in the algorithm.

5 Unanswered Questions

The Descriptive Numbers exhibit many patterns which are worthy of further investigation. I have formulated some currently unanswered questions below to spark further investigation.

1. Are there decimal string descriptive sequences of cycle length greater than 3? Are there binary string descriptive sequences of cyc length greater than 1? In general, what is the maximum possible

cycle length which can occur in a given base and how are the cycle lengths distributed in that base? Must all members of a cycle have a same length? If so, does this make it possible to design an algorithm to produce longer cycles?

2. Is there are way to predict simply the pre-cycle length of an initial condition? How are the pre-cycle lengths distributed among different initial conditions and bases? In general, what is the maximum possible pre-cycle length which can iccur in a given base?

3. Are there other isolated nodes? Is there an easy method for finding isolated nodes? In binary, is "1101001" the only funnel? In general, is the number of different funnels representable in terms of the base?

4. Is there a binary operator, $\star$, for which $T_b(x \star y) = T_b(x) \star T_b(y)$ for any strings x, y? Or, at least, is there an operator for which a similar type of decomposition can occur?

References

[1] OEIS Foundation Inc. (2011), The On-Line Encyclopedia of Integer Sequences, Sequences A005151 and A006711.

[2] V. Bonstein and A. Fraenkel, *On a Curious Property of Counting Sequences*. The American Mathematical Monthly Vol. 101, No. 6 (1994) 560-563.

[3] J. H. Conway, *The weird and wonderful chemistry of audioactive decay*, Eureka 46 (1986) 5-16.