Journal of Integer Sequences, Vol. 21 (2018),
Article 18.9.2

Iterations of the Words-to-Numbers Function

Matthew E. Coppenbarger
School of Mathematical Sciences
85 Lomb Memorial Drive
Rochester Institute of Technology
Rochester, NY 14623-5603
USA
mecsma@rit.edu

Abstract

The *Words-to-Numbers* function takes a nonnegative integer and maps it to the number of characters required to spell the number in English. For each $k = 0, 1, \ldots, 8$, we determine the minimal nonnegative integer n such that the iterates of n converge to the fixed point (through the *Words-to-Numbers* function) in exactly k iterations. Our travels take us from some simple answers, to an etymology of the names of large numbers, to Conway/Guy's established system representing the name of any integer, and finally use generating functions to arrive at a very large number.

1 Introduction and initial definitions

The idea for the problem originates in a 1965 book [3, p. 243] by recreational linguistics icon Dmitri Borgmann (1927–85) and in a 1966 article [16] by another icon, recreational mathematician Martin Gardner (1914–2010). Gardner's article appears again in a book [17, pp. 71–72, 265–267] containing reprinted Gardner articles from Scientific American along with his comments on his reader's responses. Both Borgmann and Gardner identify the only English number that requires exactly the same number of characters to spell the number as the number represents. Borgmann goes further by identifying numbers in 16 other languages with this property. Borgmann calling such numbers *truthful* and Gardner calling them *honest*.

The concept was expanded in 1972 by IBM computer scientists Gosper and Schroeppel in the unpublished HAKMEM (short for "hacker's memo") report [2, item 134] by defining a function that counts the number of characters in the English name of any representation of a number and iterating through the function until a cycle is reached. The fixed points of such a function would be truthful/honest numbers. In 1974, Kravitz [26] generalized the function and applied it to 17 other languages, identifying the fixed points and cycles generated by the iterates of elements of the domain. For purposes of this paper, we will only focus on the English version. The name of this function, *Words-to-Numbers*, was coined by Ecker in 1986 or 1987. As cited by Zerger [43], the name is given in newsletters published by Ecker ([12] or [13]), but are no longer available. Some of the material in the newsletters, which includes the *Words-to-Numbers* function, is summarized by Ecker in an article for a tribute book to Gardner [14].

The *Words-to-Numbers* function, $\mathcal{W} : \mathbb{N} \to \mathbb{Z}^+$, is the very simple process of taking a nonnegative integer n and mapping it to the positive integer representing the number of characters required to spell n in English with the conventions (a) use only a space to separate each word (no commas, hyphens, nor "`and`"s) and (b) include each space in the character count. In contrast, Sloane [39] lists sequences A005589 and A227290 that have similar definitions but do not count the space. Since it was Ecker's article [14] that motivated this paper, we use Ecker's initial intent to include the space in the count. For example, $\mathcal{W}(126) = 22$ since "`one hundred twenty six`" requires a string of 22 characters to spell. The *iteration sequence* of n is the sequence of the iterates of $\mathcal{W}$ generated by n (the *seed*). To illustrate, the iteration sequence of 126 is

$$126 \to 22 \to 10 \to 3 \to 5 \to 4 \to 4 \to \cdots , \tag{1}$$

indicating that 4 is the answer to Borgmann/Gardner's initial question. Moreover, the *height* of the iteration sequence generated by n, denoted $\mathcal{W}^{\#}(n)$, is the number of iterations of $\mathcal{W}$ that must be applied to n until the iteration sequence reaches either a fixed point or a cycle. We can see from (1) that $\mathcal{W}^{\#}(126) = 5$ since 126 requires five iterations of $\mathcal{W}$ until arriving at the fixed point.

One template for analyzing the iteration dynamics of discrete functions is given by Guy [20, problem E34]. The function used takes any positive integer and maps it to the sum of the squares of its decimal digits. Guy considers only the subset of the domain whose elements iterate to 1 (defined as the set of *happy numbers*). Among the deluge of questions proposed, he asks, for a nonnegative integer k, to determine the least nonnegative integer that converges to 1 in exactly k iterations. Solutions to problem E34 are partially answered by Grundman and Teeple [19] and Cai and Zhaou [7] and are summarized as A001273 by Sloane [39].

The main purpose of this paper is to do the same for $\mathcal{W}$, but we need to know that the fixed point of $\mathcal{W}$ identified in (1) is the only one and that there are not any cycles of length greater than one. We summarize this as

Theorem 1. *All iteration sequences through $\mathcal{W}$ converge to 4.*

A partial proof of the theorem is given in §2 and completed in §7 after choosing a numeration scheme valid for all integers.

Using $\mathcal{W}$, we can naturally partition the domain by defining, for each nonnegative integer k, the k-th *iterated set* of $\mathcal{W}$, denoted I_k, to be the subset of the domain consisting of those elements that map the fixed point in exactly k iterations. Using our previous example again, $126 \in I_5$. We let

$$\tau_k := \min(I_k)$$

denote the minimum value that maps to the fixed point in exactly k iterations. Hence the definition implies that $\tau_5 \leq 126$.

Taking inspiration from Guy, we establish the values of τ_k through $\mathcal{W}$ for as many values of k that are possible. (This iteration sequence is given as A301408 by Sloane [39].) We will see that τ_0 through τ_6 are fairly simple, τ_7 is a little more difficult, and the material needed to ascertain τ_8 encompasses over half of this article.

In addition to the sets $\mathbb{N}$ for the nonnegative integers and $\mathbb{Z}^+$ for the positive integers, we use $\mathbb{N}_b := \{0, 1, 2, \ldots, b-1\}$, where $b \in \mathbb{Z}^+$, to represent the first b nonnegative integers. Most uses of $\mathbb{N}_b$ will be to construct the base-b expansion of an integer.

2 Zero through six iterations

There are 29 words that are needed to name any nonnegative integer less than 1000 and are listed in Table 1. We use them to derive the first seven integers of minimal height.

Strings	Length
"one", "two", "six", "ten"	3
"zero", "four", "five", "nine"	4
"three", "seven", "eight", "forty", "fifty", "sixty"	5
"eleven", "twelve", "twenty", "thirty", "eighty", "ninety"	6
"fifteen", "sixteen", "seventy", "hundred"	7
"thirteen", "fourteen", "eighteen", "nineteen"	8
"seventeen"	9

Table 1: The words needed for naming all nonnegative integers less than 1000 grouped by length.

For small values in the domain, the iteration structure of $\mathcal{W}$ is chaotic and finding the number that is of minimal height does not become predictable until height 6. So our analysis begins by drawing Figure 1, a digraph representation of $\mathcal{W}$ for a particular sampling of elements in the first six iterated sets.

Since all nonnegative integers less than 29 are included in Figure 1 or listed in Table 1, then we can easily see that

Lemma 2. *The least nonnegative integers of heights 0 through 5 through $\mathcal{W}$ are, respectively, $\tau_0 = 4, \tau_1 = 0, \tau_2 = 3, \tau_3 = 1, \tau_4 = 11,$ and $\tau_5 = 23$.*

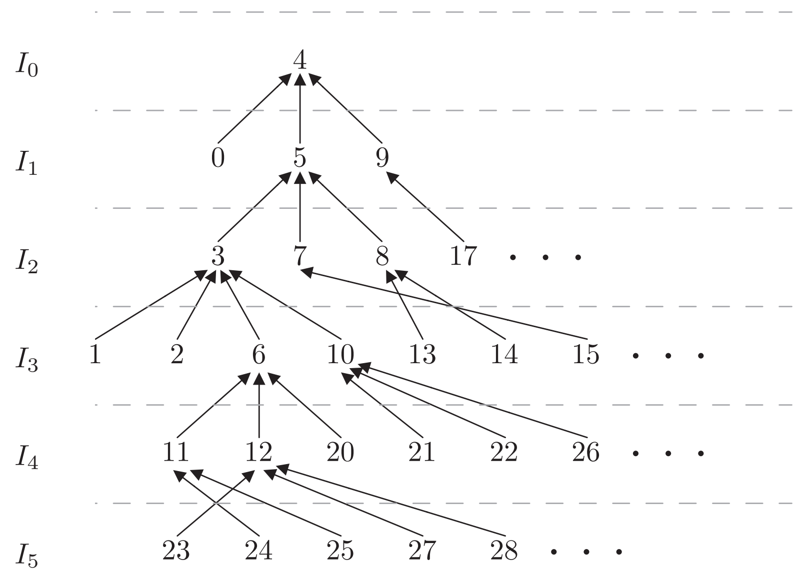

Figure 1: A partial digraph representation of the first six iterated sets of $\mathcal{W}$.

The determination of τ_6 will require a number of preliminary calculations. Given an integer n in the k-th iterated set I_k, the set of preimages of n, denoted $\mathcal{W}^{-1}(n)$, must be a subset of I_{k+1}. So we need to find the smallest integer in each of the preimage sets of elements in $I_5 = \{23, 24, 25, 27, \dots\}$, but only an exhaustive search using Table 1 will produce these numbers. We will see that similar computations must be performed for each iteration above six, so all minimal preimages less than 1000 are provided here. Formally, for $n \in \mathbb{N}$, define $T_n := \min(\mathcal{W}^{-1}(n))$ to be the minimum of all nonnegative integers that map to n through $\mathcal{W}$. The results have been compiled in Table 2.

n	T_n	n	T_n	n	T_n	n	T_n
1	*undefined*	8	13	15	101	22	121
2	*undefined*	9	17	16	104	23	124
3	1	10	21	17	103	24	123
4	0	11	24	18	111	25	173
5	3	12	23	19	115	26	323
6	11	13	73	20	113	27	373
7	15	14	≥ 1000	21	117	28+	≥ 1000

Table 2: The smallest nonnegative integer T_n that requires exactly n characters to spell.

The table indicates that the preimage of any $n \geq 28$ must consist of more than three decimal digits. Thus evaluating τ_6 reduces to knowing the preimages of the first four elements in I_5. That is,

$$\tau_6 = \min(I_6) = \min\{T_{23}, T_{24}, T_{25}, T_{27}\} = 123.$$

We summarize this result as

Lemma 3. *The least nonnegative integer of height 6 through $\mathcal{W}$ is $\tau_6 = 123$.*

Table 2 also serves an additional purpose in that it can partially answer the theorem given earlier that all iteration sequences converge to 4.

Partial proof of Theorem 1. Initially, we observe by Figure 1 that all seeds less than 11 converge to 4. We also note from Table 2 that $n < T_n$ for all n from 6 to 28. Consider a seed m such that $11 \leq m < 1000$. Since both m and $T_{\mathcal{W}(m)}$ map to $\mathcal{W}(m)$ and $T_{\mathcal{W}(m)}$ is the smallest integer that maps to $\mathcal{W}(m)$, then $m \geq T_{\mathcal{W}(m)} > \mathcal{W}(m)$, indicating that iterations of $\mathcal{W}$ are decreasing for those values of m. Thus all iteration sequences with seed $m \in \mathbb{N}_{1000}$ must invariably converge to 4. $\square$

The proof is completed in §7.

3 An etymology of the names of large numbers – and a choice

To determine τ_7, we need the name designation of three more numbers: 10^3, 10^6, and 10^9. The first two (*thousand* and *million*) are widely accepted designations, but the third is not. To explain, we step back to the last quarter of the 15th century in France to when these names were first defined and follow their usage and the evolution of their meaning.

One of the first references of the names of numbers larger than 10^6 was in 1484 by French mathematician and physician Nicolas Chuquet (c.1445–c.1488) in his manuscript *Triparty en la science des nombres* (*The Science of Numbers in Three Parts*). The beginning of the first part includes definitions of the words *billion, trillion, quadrillion, quillion, sixlion, septillion, ottillion,* and *nonillion* as the name designations of, respectively, 10^{12}, 10^{18}, 10^{24}, and so on, up to 10^{54}, thus grouping the decimal digits in *periods* of six. (These spellings are from the English translation of Chuquet's manuscript by Flegg, Hay, and Moss [15, p. 29].) *Million*, which is of Italian origin, had been in usage in the previous centuries prior to Chuquet. The idea of adding Latin cardinal numbers as prefixes to *-illion* was not new but the origin is, nonetheless, now attributed to Chuquet. In 1475, French mathematician Jehan Adam includes the words *million, bymillion,* and *trimillion* in his manuscript *Traicte en arismeticque*. It is speculated by Flegg, Hay, and Moss [15, p. 339] that Chuquet and Adam may have seen the definitions from the same source. Both mention Bertheleny de Romanis as an inspiration, but none of his work has survived.

Chuquet's manuscript was discovered by French scholar Aristide Marre (1823–1918) and not published until 1880 and 1881 [30, 31, 32], making Chuquet's contribution effectively unknown to the mathematical community for nearly four centuries. Notwithstanding, the mathematics contained in the *Triparty* was known even while unpublished because, after Chuquet's death, possession passed to fellow Frenchman Estinenne de la Roche (1470–1530). A substantial portion of the contents of Chuquet's manuscript was more widely disseminated with the publication of the arithmetic text *Larismethique nouellement compose* in 1520 by de la Roche who copied nearly verbatim large swaths of Chuquet's work, including the name designation of numbers given above.

The history of Chuquet's *Triparty* and its impact on the mathematical community are stories unto themselves. Only one copy has survived and it is the same one owned by de la Roche – his notes are in the margins. Flegg, Hay, and Moss [15] trace the manuscript from de la Roche to Marre and provide a comprehensive analysis of Chuquet's mathematics and its influence including an explanation of the conditions for which it was permissable to plagiarize in 15th/16th century France. Additional history about de la Roche's role in the legacy of Chuquet is provided by Moss [33].

In 1690, over 200 years after the *Triparty* was written, the name designations, with the same definitions as given by Chuquet/de la Roche, reappeared in the classical work *An Essay Concerning Human Understanding* [28, II-16] by the English philosopher John Locke (1632–1704). The spellings used by Locke are identical to the ones used today with one exception: Locke used *quatrillion* instead of today/Chuquet's *quadrillion*. The vast majority of subsequent authors recognized Locke's mistake and used the latter representation. As the *Essay* was Locke's most celebrated work and he presented the name designations without attribution, the *Essay* was cemented as the standard for numeration. Locke's naming convention became known as the *English method* of numeration, despite being of French origin. How the terms came to Locke has never been definitely established, but one possible link between de la Roche and Locke is provided by Moss [33]: Locke's geometry professor at Oxford, John Wallis (1616–1703), was aware of de la Roche through French monk and mathematician Jean Buéton (1492–c.1564) who wrote *Logistica, quae et arithmetica vulgo dicitur* in 1559 and cited de la Roche.

It is at this point where the definitions, and hence $\mathcal{W}$, become ill-defined. Mathematicians and educators in France in the late 18th century grouped the decimal digits in periods of three but used the same name designations as given by Locke. They defined *billion* through *nonillion* as the name designations for $10^9, 10^{12}, \ldots, 10^{30}$. As noted by historian Florian Cajori (1859–1930) [8, p. 49], France went through a revitalization of mathematics and mathematical education that originated with Swiss education reformer Johann Pestalozzi (1746–1827). Many mathematicians and educators in the United States were influenced by this reform and among the changes adopted was this new method (and referred to as the *French method* of numeration) replacing the English method. The earliest use of the French method in the United States was Patterson's 1805 edition of Dilworth's *Schoolmaster's Assistant* (as observed by Cajori [8, p. 108]) with many arithmetic textbooks rapidly transitioning to this new method between 1825 and 1830. We note that this 1805 publication date is in contrast with mathematician David E. Smith (1860–1944) who writes in one of his more popular history books [40, p. 86] that the earliest use of the French method in the American colonies is an arithmetic textbook [18] by Harvard professor Isaac Greenwood

(1702–45) in 1729. In fact, Greenwood used the English method and cited it as "Mr. Lock's [sic] Method of Numeration". Greenwood's textbook was published anonymously and was the first arithmetic textbook written in colonial America. Greenwood himself was the first Hollis Chair of Mathematics and Natural Philosophy at Harvard (and the position is still occupied today after nearly 300 years). He is also noteworthy to be the first person to teach calculus in the colonies [27]. Sadly, his academic career ended when he was censured in 1737 for drunkenness and dismissed in the next year.

Acceptance of the French method in the United States may have been popular, but it was not so universally. A noteworthy dissent was by Elijah Slack (1784–1866), a mathematician, chemist, and Presbeterian minister (and served early in his academic career as vice president of Princeton College from 1812 to 1817, a very tumultuous time in its history when the students openly revolted against the administration). After relocating to Ohio and much later in his career, Slack wrote a remarkable 1853 paper [38] where he presented a passionate defense of the English method. The paper, however, failed to resonate with the Pestalozzian reformers, perhaps because the message was lost since the first portion of the paper consisted of Slack's proclamation of the divine origin of the decimal digits. It is more likely, however, that Slack failed to dissuade the movement of the numeration community toward the French method because it was already over – the last use of the English method to appear in the United States in an arithmetic textbook was three years earlier in Gibson's revised edition of Fowler's *Youth's Assistant* (as noted by Cajori [8, p. 108]).

To complicate matters further, a movement began in parts of Europe to use the suffix *-illiard* (first proposed by French mathematician and poet Jacques Pelletier du Mans (1517–83)) for the intermediate name designations of 10^9, 10^{15}, 10^{21}, etc. (while still using the English method for the name designations for 10^{12}, 10^{18}, etc.). It was slowly adopted in many European countries and is referred to as the *Traditional European* naming designation. The French, finally realizing their earlier error in the 18th century, moved to the Traditional European naming designation in the 1960s. The English method is now called the *Traditional British* naming convention. But England, citing influence by the United States, officially switched to the French method which is now called the *U.S. and Modern British* naming convention.

Overall, there are three naming conventions for large numbers; they are given in Table 3 and make $\mathcal{W}$ ill-defined for all $n \geq 10^9$. As it is not feasible to present the remaining two iterations for all three naming conventions, we simplify by choosing only one. Since $\mathcal{W}$ has been defined as the length of a number spelled in English, we use the system prevalent with most English-speaking countries: the U.S. and Modern British naming convention.

4 Representing integers

We begin this section with a well-known result of the unique representation of an integer in a chosen base b.

Naming Convention	Name of 10^9
U.S. and Modern British (French method)	*Billion*
Traditional British (English method)	*Thousand Million*
Traditional European	*Milliard*

Table 3: The three different naming conventions and the corresponding names of 10^9.

Theorem 4. *Given $n, b \in \mathbb{N}$ with $b > 1$, there exist unique integers $\xi_1, \xi_2, \ldots, \xi_\delta \in \mathbb{N}_b = \{0, 1, \ldots, b-1\}$ such that $n = \sum_{k=1}^{\delta} \xi_k b^{k-1}$ (with $\xi_\delta \neq 0$ if $n > 0$ and $\delta = 1$ if $n = 0$).*

We shall refer to each ξ_k as a *base-b digit* of n (or, more specifically, the *k-th base-b digit* of n) with ξ_δ the *leading base-b digit* and, for $n > 0$,

$$\delta \equiv \delta_b(n) = 1 + \lfloor \log_b n \rfloor$$

(where $\lfloor \cdot \rfloor$ is the floor function) is the number of base-b digits of n.

The proof of Theorem 4 uses the division algorithm and the well-ordering principle and is omitted since it can be found in many introductory textbooks in discrete mathematics (such as Kolman, Busby, and Ross [25, p. 27]).

We call the summation representation of n in Theorem 4 the *base-b expansion* of n and use the notation

$$n = \left(\xi_\delta, \xi_{\delta-1}, \ldots, \xi_2, \xi_1\right)_b.$$

This format gives us the flexibility to use decimal digits but represent integers in any base.

To avoid the difficulty of naming numbers when the digits are grouped as one long string and, in the tradition of most countries using the U.S. and Modern British naming convention, integers consisting of five or more base-10 digits will be grouped by threes (starting on the right) with the final group on the left consisting of one to three base-10 digits. To avoid confusion with the notation of multivariable functions and the base-b expansion notation, we will not follow the tradition using a comma as a separatrix between groups, but rather a half space. Two examples are $12\,379 = (1, 2, 3, 7, 9)_{10} = (12, 379)_{1000} = (12\,379)_{10^6}$ and $(1, 0, 0)_{1000} = 10^6$.

5 Seven iterations

We return to the problem of ascertaining τ_7, the least positive integer requiring exactly seven iterations until reaching the fixed point 4 through $\mathcal{W}$. Because $\tau_6 = 123$, τ_7 will require over one hundred characters to spell so our answer must maximize the number of characters used while minimizing the number itself. From Table 2, the maximum number of characters that are necessary to spell a three decimal digit number is 27, and the smallest number with exactly that number of characters to spell is 373. Thus our initial attempt to find τ_7 will involve constructing a number consisting of "373" concatenated to itself multiple times.

Necessary before continuing, we have a definition and a modification to clarify notation. The *length* of a string s, denoted $|s|$, which is the number of characters in s. Because it may not be immediately obvious when a blank space is placed at the beginning and/or the end of a string, we will, hereafter, use the visible space symbol "␣" for a blank space.

For $n \in \mathbb{N}$, define

$$M_n := \sum_{j=0}^{n+1} 373 \cdot 1000^j = \big(\underbrace{373, 373, \ldots, 373}_{\text{repeated } n+2 \text{ times}}\big)_{1000} \tag{2}$$

to be the number constructed by writing "373" $n+2$ times. The lengths of M_0, M_1, and M_2 are, respectively,

$$\begin{aligned}
\mathcal{W}(M_0) &= \mathcal{W}(373) + |\text{“␣}\texttt{thousand}\text{␣”}| + \mathcal{W}(373) = 64, \\
\mathcal{W}(M_1) &= \mathcal{W}(373) + |\text{“␣}\texttt{million}\text{␣”}| + \mathcal{W}(M_0) = 100, \text{ and} \\
\mathcal{W}(M_2) &= \mathcal{W}(373) + |\text{“␣}\texttt{billion}\text{␣”}| + \mathcal{W}(M_1) = 136.
\end{aligned}$$

Since $\mathcal{W}(M_1) < \tau_6 < \mathcal{W}(M_2)$, then τ_7 must consist of exactly four base-1000 digits.

Somewhat anticlimactically, we forgo the thrill of discovery and present

Theorem 5. *The least nonnegative integer of height 7 through* $\mathcal{W}$ *is* $\tau_7 = 101\ 323\ 373\ 373$.

Proof. To start, we verify that $N := 101\ 323\ 373\ 373$ is of height 7. We use the value of $\mathcal{W}(M_2)$ and Table 2 to get

$$\mathcal{W}(N) = \mathcal{W}(M_2) - 2\mathcal{W}(373) + \mathcal{W}(101) + \mathcal{W}(323) = 123 = \tau_6.$$

Thus $\mathcal{W}^{\#}(N) = 7$ implying $\tau_7 \leq N$.

Next we show that all integers smaller than N must have a smaller height. Suppose $n < N$ with $\mathcal{W}^{\#}(n) = 7$. By the discussion preceding this theorem, n must consist of four base-1000 digits. Then $n = (\alpha_4, \alpha_3, \alpha_2, \alpha_1)_{1000}$ where $\alpha_j > 0$ for each $0 \leq j \leq 4$. If $\alpha_4 < 101$, then

$$\mathcal{W}(n) = \mathcal{W}(\alpha_4) + \sum\nolimits_{i=1}^{3} \mathcal{W}(\alpha_i) + 28 \leq 13 + 3 \cdot 27 + 28 = 122 < \tau_6.$$

This contradicts the assumption that $\mathcal{W}^{\#}(n) = 7$ and so $\alpha_4 = 101$. Presuming $\alpha_3 < 323$ leads to a similar conclusion, forcing $\alpha_3 = 323$. Finally, presuming either $\alpha_2 < 373$ or $\alpha_1 < 373$ each lead to similar conclusions. Thus $n = N$ and the result follows. $\square$

6 Extending the names of large numbers

The next iteration will take us to a number requiring over one hundred billion characters to spell. Not surprisingly, it is monstrous and poses a significant problem because $\mathcal{W}$ is ill-defined for most integers between τ_7 and τ_8 as there is not a consensus for the names of large numbers. Consequently, we need a numeration scheme that goes completely beyond the current accepted designations.

To define the name designations for every prefix cardinal, we start with those already established in the English language and appear in comprehensive modern dictionaries (such as Webster's dictionary [42, p. 1549]). They are summarized in Table 4. Our goal is to choose a naming convention most consistent with those names but is defined for all prefix cardinals.

Prefix cardinal	Latin name designation	Prefix cardinal	Latin name designation
0	*thousand*	11	*undecillion*
1	*million*	12	*duodecillion*
2	*billion*	13	*tredecillion*
3	*trillion*	14	*quattuordecillion*
4	*quadrillion*	15	*quindecillion*
5	*quintillion*	16	*sexdecillion*
6	*sextillion*	17	*septendecillion*
7	*septillion*	18	*octodecillion*
8	*octillion*	19	*novemdecillion*
9	*nonillion*	20	*vigintillion*
10	*decillion*	100	*centillion*

Table 4: The Latin name designations as they appear in modern dictionaries.

But when were the dictionary names beyond *nonillion* first established? We continue with where Locke left off by citing the earliest known published instances to extend the Latin name designation beyond those given by Chuquet. *Decillion*, *undecillion*, and *duodecillion* first appear in a 1788 arithmetic textbook [36, p. 19] by Nicholas Pike (1743–1819). This text is noteable as the first major arithmetic textbook in the newly created United States and is historically significant for being endorsed by George Washington (1732–99) [1, vol. 6, p. 7–8] when he reviewed it before becoming president. *Tredecillion* through *vigintillion* first appear in a revised edition [37, p. 19] of Pike's text published posthumously in 1822. The word *centillion* first appears in an 1834 collection of short stories [41, p. 118], but it is not until 1853 in the same article by Slack [38] referenced earlier where *centillion* is given with its proper definition.

There are only a handful of published sources that provide names filling the gap in Table 4 (that is, the name designation corresponding to the prefix cardinals $m = 21$ through $m = 99$). The first is the article by Slack [38] again. He provides the name designations corresponding to prefix cardinals up to $m = 27$ and for each of the multiples of 10 from $m = 30$ to $m = 100$. Similar presentations follow by Heath [21, p. 17] in 1856, Holbrook [23, p. 306] and Loomis [29, p. 18]) in 1859, and much later by Borgmann [3, p. 223] (the same book defining truthful numbers) in 1965.

Beyond centillion, the references are remote. This first is by Slack [38] again whose last entry in the article is his name designation corresponding to $m = 1000$ which he calls *millillion*. Holbrook [23, p. 306] provides names corresponding to the prefix cardinals $m = 101$, 200, and 1000 (using the same name for $m = 1000$ as did Slack). Far surpassing all others, Henkle [22] in 1860 (and reprinted with subsequent corrections by Brooks [6, p. 100, 571] in 1876) presents a system with over 100 additional name designations but constructed utilizing a mixture of Latin cardinals and ordinals (although he does use the same name designation for $m = 1000$ as given by Slack and Holbrook). The last designation by Henkle is for $m = 1\,000\,000$ which he calls *milli-millillion*. The other designations, however, are completely unsystematic and not presentable in any compact way. An attempt was made in a series of articles over a 10 year span in *Word Ways: The Journal of Recreational Linguistics* [4, 5, 9, 10, 35] to modify the extended names in Henkle/Brooks and construct a more consistent system. Their overall result, however, is only marginally better.

Conway and Guy's *The Book of Numbers* [11, p. 14–15] provides the most notable and complete naming extension that is consistent with Chuquet's initial intent and we will use it to calculate τ_8. For easier reference, we christen it the *CGW system* for the Latin name designation of large numbers. (The "W" in the designation is because the authors credit Allan Wechsler as a contributor to their system.)

We review some needed terminology for strings and formalize a few definitions before presenting the CGW system (given in Definition 6 below). The string consisting of no characters, denoted λ, is the *empty string* (with length $|\lambda| = 0$). Given two strings s and t, we denote the *concatenation* as st (with the usual convention that $\lambda s = s\lambda = s$). Given a collection of k strings $s_1, s_2, \ldots, s_k$, we use the product notation $\prod_{i=1}^{k} s_i$ to define the concatenation recursively as s_1 if $k = 1$ and $s_k \prod_{i=1}^{k-1} s_i$ otherwise (so that the resultant string $s_k s_{k-1} \ldots s_2 s_1$ is ordered from right to left). Given a string s and a nonnegative integer k, we define the k-th *power* of s, denoted s^k as λ if $k = 0$ and $s^k := \prod_{i=1}^{k} s$ otherwise (that is, s^k consists of s written k times with $s^0 = \lambda$). For a nonnegative integer m, define $\mathcal{L}^*(m)$ to be the *Latin name designation* of a one to three digit positive number followed by $3(m + 1)$ zeros (so $\mathcal{L}^*(1) =$ "`million`"). Accordingly, m is called the *prefix cardinal* associated to a given Latin name designation. In a similar manner, we define $\mathcal{W}^*$ to be the name designation of a nonnegative integer (so $\mathcal{W}^*(126) =$ "`one_hundred_twenty_six`"). The lengths of these strings are denoted by the functions $\mathcal{L} \equiv |\mathcal{L}^*|$ and $\mathcal{W} \equiv |\mathcal{W}^*|$.

The following definition is a formal representation to the very colloquial (but equivalent) version given by Conway/Guy.

Definition 6 (*The CGW system*)**.** The Latin name designation associated to the prefix cardinal m is

$$\mathcal{L}^*(m) := \begin{cases} \text{“}\texttt{thousand}\text{”}, & \text{if } m = 0; \\ \left(\prod_{j=1}^{\rho} \Gamma_{\omega_j} \text{“}\texttt{illi}\text{”}\right) \text{“}\texttt{on}\text{”}, & \text{if } m > 0 \end{cases} \tag{3}$$

where, for $m > 0$, the base-1000 expansion is $m = (\omega_\rho, \ldots, \omega_2, \omega_1)_{1000}$ with $\rho = 1 + \lfloor \log_{1000} m \rfloor$. We refer to each Γ_η as a *base string*. For each $\eta \in \mathbb{N}_{1000}$, there are associated unique integers $\eta_1, \eta_2, \eta_3 \in \mathbb{N}_{10}$ such that $\eta = 100\eta_3 + 10\eta_2 + \eta_1$. The base string is

$$\Gamma_\eta := \begin{cases} \text{“}\mathtt{n}\text{”}, & \text{if } \eta = 0; \\ \mathfrak{q}_\eta, & \text{if } 0 < \eta < 10; \\ \mathfrak{A}_{\eta_1}\mu_\eta\mathfrak{B}_{\eta_2}\nu_\eta\mathfrak{C}_{\eta_3}, & \text{if } \eta \geq 10 \end{cases} \tag{4}$$

where the strings $\mathfrak{q}_i$, $\mathfrak{A}_i$, $\mathfrak{B}_i$, and $\mathfrak{C}_i$, are given by Table 5. The connective components μ_η and ν_η each consist of no more than a single character and are given by

$$\mu_\eta := \begin{cases} \Phi_{\eta_1,\eta_3}, & \text{if } \eta_2 = 0; \\ \Psi_{\eta_1,\eta_2}, & \text{if } \eta_2 > 0 \end{cases} \quad \text{and} \quad \nu_\eta := \begin{cases} \lambda, & \text{if } \eta_2 = 0 \text{ or } \eta_3 = 0; \\ \text{“}\mathtt{i}\text{”}, & \text{if } \eta_2 = 1, 2 \text{ and } \eta_3 > 0; \\ \text{“}\mathtt{a}\text{”}, & \text{if } \eta_2 > 2 \text{ and } \eta_3 > 0 \end{cases} \tag{5}$$

where the characters for Φ and Ψ are given in, respectively, Tables 6 and 7.

i	$\mathfrak{q}_i$ **(Chuquet)**	$\mathfrak{A}_i$ **(Units)**	$\mathfrak{B}_i$ **(Tens)**	$\mathfrak{C}_i$ **(Hundreds)**
0	λ	λ	λ	λ
1	“m”	“un”	“dec”	“cent”
2	“b”	“duo”	“vigint”	“ducent”
3	“tr”	“tre”	“trigint”	“trecent”
4	“quadr”	“quattuor”	“quadragint”	“quadringent”
5	“quint”	“quinqua”	“quinquagint”	“quingent”
6	“sext”	“se”	“sexagint”	“sescent”
7	“sept”	“septe”	“septuagint”	“septingent”
8	“oct”	“octo”	“octogint”	“octingent”
9	“non”	“nove”	“nonagint”	“nongent”

Table 5: The components of the base strings of the Chuquet names and the CGW system.

We illustrate the definition by looking at a few simple examples. If $m = 1 = (1)_{1000}$, then $\rho = 1$ and

$$\mathcal{L}^*(1) = \left(\prod_{j=1}^{\rho} \Gamma_{\omega_j} \text{“}\mathtt{illi}\text{”}\right) \text{“}\mathtt{on}\text{”} = (\Gamma_1 \text{“}\mathtt{illi}\text{”})\, \text{“}\mathtt{on}\text{”} = \mathfrak{q}_1 \text{“}\mathtt{illion}\text{”} = \text{“}\mathtt{million}\text{”}$$

as expected. We note that $\mathcal{L}^*(1000) = \Gamma_1 \text{“}\mathtt{illi}\text{”} \Gamma_0 \text{“}\mathtt{illion}\text{”} = \text{“}\mathtt{millinillion}\text{”}$ demonstrates that the use of “n” as a placeholder for some larger prefix cardinals (making this slightly inconsistent with the designations by Slack and Henkle). A longer Latin name designation sampling of the rules in the CGW system is

$$\begin{aligned} \mathcal{L}^*(27\,002\,103\,300) &= \Gamma_{27}\text{“}\mathtt{illi}\text{”}\Gamma_2\text{“}\mathtt{illi}\text{”}\Gamma_{103}\text{“}\mathtt{illi}\text{”}\Gamma_{300}\text{“}\mathtt{illion}\text{”} \\ &= \mathfrak{A}_7\text{“}\mathtt{m}\text{”}\mathfrak{B}_2\lambda\lambda\text{“}\mathtt{illi}\text{”}\mathfrak{q}_2\text{“}\mathtt{illi}\text{”}\mathfrak{A}_3\text{“}\mathtt{s}\text{”}\lambda\lambda\mathfrak{C}_1\text{“}\mathtt{illi}\text{”}\lambda\lambda\lambda\lambda\mathfrak{C}_3\text{“}\mathtt{illion}\text{”} \\ &= \text{“}\mathtt{septemvigintillibillitrescentillitrecentillion}\text{”}. \end{aligned}$$

η_1	η_3 0	1	2	3	4	5	6	7	8	9
0	λ	λ	λ	λ	λ	λ	λ	λ	λ	λ
1	λ	λ	λ	λ	λ	λ	λ	λ	λ	λ
2	λ	λ	λ	λ	λ	λ	λ	λ	λ	λ
3	λ	"s"	λ	"s"	"s"	"s"	λ	λ	"s"	λ
4	λ	λ	λ	λ	λ	λ	λ	λ	λ	λ
5	λ	λ	λ	λ	λ	λ	λ	λ	λ	λ
6	λ	"x"	λ	"s"	"s"	"s"	λ	λ	"x"	λ
7	λ	"n"	"n"	"n"	"n"	"n"	"n"	"n"	"m"	λ
8	λ	λ	λ	λ	λ	λ	λ	λ	λ	λ
9	λ	"n"	"n"	"n"	"n"	"n"	"n"	"n"	"m"	λ

Table 6: Components of Φ_{η_1,η_3} for the CGW system (where $\eta_2 = 0$).

A more explicit example to consider is the GOOGOL sequence of numbers. It is defined recursively as $g_0 := 100$ and $g_n := 10^{g_{n-1}}$ for $n \in \mathbb{Z}^+$. The second term of the sequence, $g_1 = 10^{100}$, is the GOOGOL (first defined by Kasner and Newman [24, p. 20]). Since $g_1 = 10^{100} = 10 \cdot 1000^{1+32}$, then

$$\mathcal{W}^*(g_1) = \mathcal{W}^*(10)\text{“␣”}\mathcal{L}^*(32) = \text{“}\texttt{ten␣}\text{”}\mathfrak{A}_2\lambda\mathfrak{B}_3\lambda\lambda\text{“}\texttt{illion}\text{”} = \text{“}\texttt{ten␣duotrigintillion}\text{”}.$$

The third term of the sequence, g_2, is the GOOGOLPLEX. Since

$$\begin{aligned} g_2 = 10^{g_1} = 10^{1+\sum_{i=1}^{100} 9\cdot 10^{i-1}} &= 10 \cdot 10^{3\sum_{i=1}^{100} 3\cdot 10^{i-1}} \\ &= 10 \cdot 1000^{1+\left(3\cdot 1000^{33}+\sum_{j=2}^{33} 333\cdot 1000^{j-1}+332\right)} \\ &= 10 \cdot 1000^{1+m_{g_2}}, \end{aligned}$$

where $m_{g_2} = (3, 333, 333, \ldots, 333, 332)_{1000}$ is the associated prefix cardinal (and consists of 34 base-1000 digits), then the Latin name designation is

$$\begin{aligned} &\mathcal{W}^*(g_2) \\ &= \mathcal{W}^*(10)\text{“␣”}\mathcal{L}^*(m_{g_2}) \\ &= \text{“}\texttt{ten␣}\text{”}\Gamma_3\text{“}\texttt{illi}\text{”}(\Gamma_{333}\text{“}\texttt{illi}\text{”})^{32}\Gamma_{332}\text{“}\texttt{illion}\text{”} \\ &= \text{“}\texttt{ten␣}\text{”}\mathfrak{q}_3\text{“}\texttt{illi}\text{”}(\mathfrak{A}_3\text{“}\texttt{s}\text{”}\mathfrak{B}_3\text{“}\texttt{a}\text{”}\mathfrak{C}_3\text{“}\texttt{illi}\text{”})^{32}\mathfrak{A}_2\lambda\mathfrak{B}_3\text{“}\texttt{a}\text{”}\mathfrak{C}_3\text{“}\texttt{illion}\text{”} \\ &= \text{“}\texttt{ten␣trilli}\text{”}\text{“}\texttt{trestrigintatrecentilli}\text{”}^{32}\text{“}\texttt{duotrigintatrecentillion}\text{”} \end{aligned}$$

and requires $\mathcal{W}(g_2) = 10 + 23 \cdot 32 + 24 = 770$ characters to spell.

One criticism of the CGW system is its inconsistency with three of the dictionary definitions in Table 4. They are $\mathcal{L}^*(15) =$ "quinquadecillion", $\mathcal{L}^*(16) =$ "sedecillion", and $\mathcal{L}^*(19) =$ "novendecillion". Although it was highly desirous to have chosen a system that

η_1	η_2 1	2	3	4	5	6	7	8	9
0	λ	λ	λ	λ	λ	λ	λ	λ	λ
1	λ	λ	λ	λ	λ	λ	λ	λ	λ
2	λ	λ	λ	λ	λ	λ	λ	λ	λ
3	λ	“s”	“s”	“s”	“s”	λ	λ	“s”	λ
4	λ	λ	λ	λ	λ	λ	λ	λ	λ
5	λ	λ	λ	λ	λ	λ	λ	λ	λ
6	λ	“s”	“s”	“s”	“s”	λ	λ	“x”	λ
7	“n”	“m”	“n”	“n”	“n”	“n”	“n”	“m”	λ
8	λ	λ	λ	λ	λ	λ	λ	λ	λ
9	“n”	“m”	“n”	“n”	“n”	“n”	“n”	“m”	λ

Table 7: Components of Ψ_{η_1,η_2} for the CGW system (where $\eta_2 > 0$ and $\eta_3 \in \mathbb{N}_{10}$).

corresponds with the dictionary definitions, the designer's choice was deliberate knowing that it did conflict and instead choosing a spelling they determined more consistent with the Latin spellings of numbers. We therefore utilize the CWG system exactly as it was intended by Conway/Guy.

7 The Completion of the Proof of Theorem 1

With the name designations of all nonnegative integers defined, we can now provide the (obvious!) proof that all iteration sequences converge to the fixed point (proposed in §1 and partially proven in §2). The proof utilizes a bound on the longest possible base string Γ_η (with $\eta \in \mathbb{N}_{1000}$) that can be constructed in Definition 6. A little experimenting with the definition and Table 5 can easily produce the upper bound of $|\Gamma_{454}| = |\mathfrak{A}_4\lambda\mathfrak{B}_5\text{“a”}\mathfrak{C}_4| = 31$. As in the partial proof, this remaining part is also based on the counting of the length of M_n given by (2). For $n \in \mathbb{N}$, we have

$$\begin{aligned}\mathcal{W}(M_n) &= \sum\nolimits_{i=0}^{n} \Big(\mathcal{W}(373) + |\text{“␣”}| + \mathcal{L}(i) + |\text{“␣”}|\Big) + \mathcal{W}(373)\\ &= 29n + 56 + \sum\nolimits_{i=0}^{n} \mathcal{L}(i) = 29n + 64 + \sum\nolimits_{i=1}^{n} \mathcal{L}(i)\end{aligned} \tag{6}$$

where we take the convention that the summand is zero if the upper bound of summation is less than the lower bound.

Theorem 1. *All iteration sequences through $\mathcal{W}$ converge to 4.*

Completion of the proof. As in the partial proof in §2, we show that $\mathcal{W}(n) < n$ for all $n \geq 10^3$.

If $10^3 \le n < 10^6$, then $\mathcal{W}(n) \le \mathcal{W}(M_0) = 64 < n$.

For a given $n \ge 10^6$, let δ be the number of base-1000 digits of n. Then

$$\delta = 1 + \lfloor \log_{1000} n \rfloor \le 1 + \log_{1000} n$$

and hence $\delta \ge 3$. Let κ be the number of base-1000 digits of $\delta - 2$. Then

$$\kappa = 1 + \lfloor \log_{1000}(\delta - 2) \rfloor \le 1 + \log_{1000}(\delta - 2).$$

We can easily verify that, for any $n \ge 10^6$, $29\delta + 6 < \frac{1}{2}n$, $\delta - 2 < \frac{1}{2}\sqrt{n}$, and $35\kappa + 2 < \sqrt{n}$. Then, using (6),

$$\begin{aligned} \mathcal{W}(n) \le \mathcal{W}(M_{\delta-2}) &= 29(\delta - 2) + 64 + \sum\nolimits_{i=1}^{\delta-2} \mathcal{L}(i) \\ &\le 29\delta + 6 + (\delta - 2)\left(\sum\nolimits_{j=1}^{\kappa} |\Gamma_{454}\text{“}\mathtt{illi}\text{”}| + |\text{“}\mathtt{on}\text{”}|\right) \\ &= 29\delta + 6 + (\delta - 2)(35\kappa + 2) \le \tfrac{1}{2}n + \tfrac{1}{2}\sqrt{n}\sqrt{n} = n \end{aligned}$$

and hence $\mathcal{W}(n) < n$ for all $n \ge 10^6$.

Therefore, using this and the partial proof given in §2, all iteration sequences with seed $n \in \mathbb{N}$ must converge to 4. □

The proof outlines that the growth rate of $\mathcal{W}$ using the CGW system is $\mathcal{O}\big(\log n \cdot \log\log n\big)$. It is not always true that any naming convention will have this growth rate. For example, the one provided by Noll [34] is a slightly worse $\mathcal{O}\big(\log n \cdot (\log\log n)^2\big)$ despite it having a very similar appearance to Conway/Guy.

8 Counting the contributions of μ_η in the CGW system

The process of establishing τ_8 will require us to count all of the string components in Definition 6. Of these, the hardest part will be the contribution of the characters from μ_n (specifically, Φ and Ψ from Tables 6 and 7). We expedite the computations by grouping together in the tables neighboring characters by rows. For $p, q \in \mathbb{Z}^+$ and $d \in \mathbb{N}$, let

$$S_d(p, q) := \{d + 10i + 100j : i \in \mathbb{N}_p, j \in \mathbb{N}_q\} \tag{7}$$

be the set of pq numbers constructed by starting at d and increasing, in base-10, the tens digits by 0 through $p-1$ and increasing the hundreds digit by 0 through $q-1$. (We presume that p and q are small enough so that neither the tens digit nor the hundreds digit go above 9.) To illustrate, an example is $S_{270}(2, 3) = \{270, 280, 370, 380, 470, 480\}$. We partition the contributions of Φ and Ψ into six sets.

- In Φ (Table 6), the location of the three contiguous “`s`”s in the third and sixth rows are represented by, respectively, $S_{303}(1, 3)$ and $S_{306}(1, 3)$. Let D_1 be the union of these two sets.

- In Φ, the location of the eight contiguous characters (one "m" and seven "n"s) in the seventh and ninth rows are represented by, respectively, $S_{107}(1,8)$ and $S_{109}(1,8)$. Let D_2 be the union of these two sets.

- The location of the four remaining characters in Φ (two "s"s in the third row and two "x"s in the sixth row) are represented by $S_{103}(1,1)$, $S_{106}(1,1)$, $S_{803}(1,1)$, and $S_{806}(1,1)$. Let D_3 be the union of these four sets.

- In Ψ (Table 7), the location of the four contiguous "s"s in the third and sixth rows are represented by, respectively, $S_{23}(4,10)$ and $S_{26}(4,10)$. Let D_4 be the union of these two sets.

- In Ψ, the location of the eight contiguous characters (one "m" and seven "n"s) in the seventh and ninth rows are represented by, respectively, $S_{17}(8,10)$ and $S_{19}(8,10)$. Let D_5 be the union of these two sets.

- The location of the two remaining characters in Ψ (the "s" and the "x" in the eighth column) are represented by $S_{83}(1,10)$ and $S_{86}(1,10)$. Let D_6 be the union of these two sets.

Thus, for $\omega \in \mathbb{N}_{1000}$, the length of μ_ω is 1 if $\omega \in \bigcup_{i=1}^{6} D_i$ and 0 otherwise.

9 Generating functions

To get an exact value of τ_8, we follow the same format used in §5 to find τ_7. This will require us to find a systematic way to count the number of characters in $\mathcal{W}(M_n)$, where M_n is given by (2).

Since there are a large number of components where the length needs to be determined, we use generating functions because they are compact and can be easily combined into a single answer.

Lemma 7 (*Generating functions for base-b digit counting*)**.** *Let $b \in \mathbb{N}$ with $b > 1$ be a given base and $k \in \mathbb{N}$.*

- *Let $P_{b,k}(x)$ be the generating function whose coefficient of x^n is the number of integers in the set $\{1, 2, \ldots, n\}$ consisting of $k+1$ or more base-b digits. Then*

$$P_{b,k}(x) = \frac{x^{b^k}}{(1-x)^2}. \tag{8}$$

- *For $d \in \mathbb{N}_b - \{0\}$ a nonzero base-b digit, let $Q_{b,d,k}(x)$ be the generating function whose coefficient of x^n is the number of integers in the set $\{1, 2, \ldots, n\}$ such that the $(k+1)$-th base-b digit is d. Then*

$$Q_{b,d,k}(x) = \frac{x^{b^k d}(1 - x^{b^k})}{(1-x)^2 \left(1 - x^{b^{k+1}}\right)}. \tag{9}$$

Proof. Only four identities are needed. The first two are the finite and infinite geometric series. They are

$$\frac{1-x^{r+1}}{1-x} = 1 + x + x^2 + \cdots + x^r \text{ and } \frac{1}{1-x} = 1 + x + x^2 + \cdots$$

(presuming $r \in \mathbb{N}$). Differentiating each gives

$$\frac{1-(r+1)x^r + rx^{r+1}}{(1-x)^2} = 1 + 2x + 3x^2 + \cdots + rx^{r-1}$$

(presuming $r \in \mathbb{Z}^+$) and

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \cdots .$$

Since b^k is the smallest integer consisting of $k+1$ base-b digits, then

$$P_{b,k}(x) = x^{b^k} + 2x^{b^k+1} + 3x^{b^k+2} + \cdots = x^{b^k}\left(1 + 2x + 3x^2 + \cdots\right) = \frac{x^{b^k}}{(1-x)^2}.$$

To demonstrate $Q_{b,d,k}(x)$, we note that the first occurrence of d in the $(k+1)$-th position is the integer $b^k d$. The occurrence of d in that position continues until $b^k(d+1)-1$ (and each of these numbers consist of exactly $k+1$ base-b digits). The next occurrence in the $(k+1)$-th position is at $b^{k+1} + b^k d$ (consisting of $k+2$ base-b digits) and continues until $b^{k+1}+b^k(d+1)-1$. In general, for $j \in \mathbb{N}$, all integers from $j \cdot b^{k+1} + b^k d$ to $j \cdot b^{k+1} + b^k(d+1) - 1$ have the digit d in the $(k+1)$-th position. Thus

$$\begin{aligned}
&Q_{b,d,k}(x) \\
&\quad = x^{b^k d}\left(1 + 2x + 3x^2 + \cdots + b^k x^{b^k-1}\right) + b^k x^{b^k(d+1)}\left(1 + x + x^2 + \cdots\right) \\
&\qquad + x^{b^{k+1}+b^k d}\left(1 + 2x + 3x^2 + \cdots + b^k x^{b^k-1}\right) + b^k x^{b^{k+1}+b^k(d+1)}\left(1 + x + x^2 + \cdots\right) \\
&\qquad + x^{2b^{k+1}+b^k d}\left(1 + 2x + 3x^2 + \cdots + b^k x^{b^k-1}\right) + b^k x^{2b^{k+1}+b^k(d+1)}\left(1 + x + x^2 + \cdots\right) \\
&\qquad + \cdots \\
&\quad = x^{b^k d}\left(1 + x^{b^{k+1}} + x^{2b^{k+1}} + \cdots\right)\left(1 + 2x + 3x^2 + \cdots + b^k x^{b^k-1}\right) \\
&\qquad + b^k x^{b^k(d+1)}\left(1 + x^{b^{k+1}} + x^{2b^{k+1}} + \cdots\right)\left(1 + x + x^2 + \cdots\right) \\
&\quad = x^{b^k d}\left(1 + x^{b^{k+1}} + (x^{b^{k+1}})^2 + \cdots\right)\left(\frac{1-(b^k+1)x^{b^k} + b^k x^{b^k+1}}{(1-x)^2} + b^k x^{b^k}\frac{1}{1-x}\right) \\
&\quad = \frac{x^{b^k d}(1-x^{b^k})}{(1-x)^2\left(1-x^{b^{k+1}}\right)}.
\end{aligned}$$

$\square$

Since (9) includes integers whose leading digit is d (with d nonzero), the lemma does not apply to cases where $d = 0$. So we present the following to specifically count those cases (but the corollary applies to the nonzero digits as well).

Corollary 8. *For $b \in \mathbb{N}$ with $b > 1$ a given base, $k \in \mathbb{N}$, and $d \in \mathbb{N}_b$, let $Q^{\star}_{b,d,k}(x)$ be the generating function whose coefficient of x^n is the number of occurrences of d in the set $\{1, 2, \ldots, n\}$ as the $(k+1)$-th base-b digit in a number consisting of $k+2$ or more base-b digits. Then*

$$Q^{\star}_{b,d,k}(x) = x^{b^{k+1}} Q_{b,d,k}(x).$$

Proof. Following the same format used for $Q_{b,d,k}(x)$ in Lemma 7 except start at $b^{k+1} + b^k d$ instead of $b^k d$. ∎

For counting contiguous blocks of characters (needed for counting the contribution of characters from μ_n in Definition 6 and §8), we have the following lemma.

Lemma 9. *Given the set $S_d(p, q)$ (with $d \neq 0$) defined in (7), let $R_{S_d(p,q),k}(x)$ be the generating function whose coefficient of x^n is the number of occurrences of any element of $S_d(p, q)$ in the set $\{1, 2, \ldots, n\}$ as the $(k+1)$-th base-1000 digit. Then*

$$R_{S_d(p,q),k}(x) = \frac{y_k^d(1 - y_k)(1 - y_k^{10p})(1 - y_k^{100q})}{(1 - x)^2(1 - y_k^{10})(1 - y_k^{100})(1 - y_k^{1000})}$$

where $y_k \equiv y_k(x) = x^{1000^k}$.

Proof. Using Lemma 7 and the formula for the sum of a finite geometric series, we have

$$\begin{aligned}
R_{S_d(p,q),k}(x) = \sum_{\omega \in S_d(p,q)} Q_{1000,\omega,k}(x) &= \frac{1 - y_k}{(1 - x)^2(1 - y_k^{1000})} \sum_{\omega \in S_d(p,q)} y_k^{\omega} \\
&= \frac{1 - y_k}{(1 - x)^2(1 - y_k^{1000})} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} y_k^{d+10i+100j} \\
&= \frac{y_k^d(1 - y_k)}{(1 - x)^2(1 - y_k^{1000})} \sum_{i=0}^{p-1} \left(y_k^{10}\right)^i \sum_{j=0}^{q-1} \left(y_k^{100}\right)^j \\
&= \frac{y_k^d(1 - y_k)}{(1 - x)^2(1 - y_k^{1000})} \cdot \frac{1 - y_k^{10p}}{1 - y_k^{10}} \cdot \frac{1 - y_k^{100q}}{1 - y_k^{100}}.
\end{aligned}$$

∎

10 Eight iterations

The two theorems in this section are the culmination of the previous four sections.

The final element needed to present the theorems is to convert the lengths of the strings in Table 5 to generating functions. We represent the lengths as four vectors:

$$\begin{aligned}
\mathbf{v}_0 &= \langle 0, 1, 1, 2, 5, 5, 4, 4, 3, 3 \rangle, \\
\mathbf{v}_1 &= \langle 0, 2, 3, 3, 8, 7, 2, 5, 4, 4 \rangle, \\
\mathbf{v}_2 &= \langle 0, 3, 6, 7, 10, 11, 8, 10, 8, 8 \rangle, \text{ and} \\
\mathbf{v}_3 &= \langle 0, 4, 6, 7, 11, 8, 7, 10, 9, 7 \rangle.
\end{aligned}$$

The i-th component of each vector is, respectively, the lengths of the strings $\mathfrak{q}_i$, $\mathfrak{A}_i$, $\mathfrak{B}_i$, and $\mathfrak{C}_i$. For each $j \in \mathbb{N}_4$, define the polynomial $V_j(z) := \mathbf{v}_j \cdot \langle 1, z, \ldots, z^9 \rangle$.

Theorem 10. *Let $F(x)$ be the generating function whose coefficient of x^n is $\mathcal{W}(M_n)$ where M_n is given by (2). Then*

$$F(x) = \frac{64}{1-x} + \frac{1}{(1-x)^2} \left(31x + \sum_{k=0}^{\infty} \left(4y_k + y_k^{100} + H(y_k) \right) \right)$$

where $y_k \equiv y_k(x) = x^{1000^k}$,

$$\begin{aligned}
H(z) &= \frac{(1-z)V_1(z)}{1-z^{10}} + \frac{G_1(z)}{1-z^{100}} + \frac{G_2(z)}{1-z^{1000}} + \frac{1-z}{1-z^{100}} \left(\frac{G_3(z)}{1-z^{10}} + \frac{G_4(z)}{1-z^{1000}} \right), \\
G_1(z) &= \left(1-z^{10}\right)\left(V_2(z^{10}) - z^{100}\right) + (1-z)\left(z^{83} + z^{86}\right), \\
G_2(z) &= (1-z^{100})V_3(z^{100}) + z^{1000}\left(z^{100} - z^{10}\right) \\
&\qquad + (1-z)\left(z^{103} + z^{106} + z^{803} + z^{806} + z^{1000} + V_0(z) - V_1(z)\right), \\
G_3(z) &= \left(z^{23} + z^{26}\right)(1-z^{40}) + \left(z^{17} + z^{19}\right)\left(1 - z^{80}\right),
\end{aligned}$$

and

$$G_4(z) = \left(z^{303} + z^{306}\right)(1-z^{300}) + \left(z^{107} + z^{109}\right)\left(1 - z^{800}\right).$$

Proof. Using (6) and the geometric series formulas given in the proof of Lemma 7, we have

$$\begin{aligned}
F(x) = \sum_{n=0}^{\infty} \mathcal{W}(M_n) x^n &= \sum_{n=0}^{\infty} \left(29n + 64 + \sum_{i=1}^{n} \mathcal{L}(i) \right) x^n \\
&= 64 \sum_{n=0}^{\infty} x^n + 29 \sum_{n=0}^{\infty} n x^n + \sum_{n=0}^{\infty} \left(\sum_{i=1}^{n} \mathcal{L}(i) \right) x^n \\
&= \frac{64}{1-x} + \frac{29x}{(1-x)^2} + \sum_{n=1}^{\infty} \sum_{i=1}^{n} \mathcal{L}(i) x^n. \qquad (10)
\end{aligned}$$

We will count the double sum $\sum_{n=1}^{\infty} \sum_{i=1}^{n} \mathcal{L}(i) x^n$ in terms of each string component appearing in Definition 6 with n representing the Latin prefix cardinal. The counting is done with regard to the base-10 or base-1000 digits, wherever appropriate. Each generating function is evaluated using Lemma 7, Corollary 8, or Lemma 9, as needed.

In (3), the “on” string appears once for every positive prefix cardinal, so the generating function to count the number of characters is

$$|\text{“}\mathtt{on}\text{”}|P_{b,0}(x) = \frac{2x}{(1-x)^2} \tag{11}$$

(any choice of $b > 1$ will work). The “illi” string occurs once for every base-1000 digit. So the generating function for the $(k+1)$-th base-1000 digit is

$$|\text{“}\mathtt{illi}\text{”}|P_{1000,k}(x) = \frac{4y_k}{(1-x)^2}. \tag{12}$$

In (4), “n” occurs every time the $(k+1)$-th base-1000 digit is 0, so the generating function is

$$|\text{“}\mathtt{n}\text{”}|Q^{\star}_{1000,0,k}(x) = \frac{y_k^{1000}(1-y_k)}{(1-x)^2(1-y_k^{1000})}. \tag{13}$$

We next count the strings appearing in Table 5. For the Chuquet strings $\mathfrak{q}_i$, the generating function for the $(k+1)$-th base-1000 digit is

$$\sum_{d=0}^{9} |\mathfrak{q}_d| Q_{1000,d,k}(x) = \frac{(1-y_k)V_0(y_k)}{(1-x)^2(1-y_k^{1000})}. \tag{14}$$

For $\mathfrak{A}_i$, the unit component of the base strings (used only when the tens or the hundreds component is nonzero), a generating function for all strings for the $(k+1)$-th base-1000 digit is

$$\sum_{d=0}^{9} |\mathfrak{A}_d| \big(Q_{10,d,3k}(x) - Q_{1000,d,k}(x)\big) = \frac{(1-y_k)V_1(y_k)}{(1-x)^2}\left(\frac{1}{1-y_k^{10}} - \frac{1}{1-y_k^{1000}}\right). \tag{15}$$

For $\mathfrak{B}_i$, the tens component of the base strings, a generating function for all strings for the $(k+1)$-th base-1000 digit is

$$\sum_{d=0}^{9} |\mathfrak{B}_d| Q_{10,d,3k+1}(x) = \frac{(1-y_k^{10})V_2(y_k^{10})}{(1-x)^2(1-y_k^{100})}. \tag{16}$$

For $\mathfrak{C}_i$, the hundreds component of the base strings, a generating function for all strings for the $(k+1)$-th base-1000 digit is

$$\sum_{d=0}^{9} |\mathfrak{C}_d| Q_{10,d,3k+2}(x) = \frac{(1-y_k^{100})V_3(y_k^{100})}{(1-x)^2(1-y_k^{1000})}. \tag{17}$$

In (5), the length of ν_η is 1 if both the tens digit and hundreds digit of η are nonzero and 0 otherwise. We count all the base-1000 digits, throw out the cases where the hundreds digits is 0 and the tens digits is 0 and then cases that were thrown out twice have to be put back in. The generating function is

$$P_{10,3k+2}(x) - Q^\star_{10,0,3k+1}(x) - Q^\star_{10,0,3k+2}(x) + \sum_{d=0}^{9} Q^\star_{1000,d,k}(x)$$
$$= \frac{1}{(1-x)^2}\left(y_k^{100} - \frac{y_k^{100}(1-y_k^{10})}{1-y_k^{100}} + \frac{y_k^{1000}(y_k^{100}-y_k^{10})}{1-y_k^{1000}}\right). \tag{18}$$

For μ_η, we use Lemma 9 and the decomposition by sets D_1 through D_6 as presented in §8. The generating functions are

$$\sum_{d\in D_1} R_{S_d(1,3)}(x) = \frac{(1-y_k)(y_k^{303}+y_k^{306})(1-y_k^{300})}{(1-x)^2(1-y_k^{100})(1-y_k^{1000})}, \tag{19}$$

$$\sum_{d\in D_2} R_{S_d(1,8)}(x) = \frac{(1-y_k)(y_k^{107}+y_k^{109})(1-y_k^{800})}{(1-x)^2(1-y_k^{100})(1-y_k^{1000})}, \tag{20}$$

$$\sum_{d\in D_3} R_{S_d(1,1)}(x) = \frac{(1-y_k)(y_k^{103}+y_k^{106}+y_k^{803}+y_k^{806})}{(1-x)^2(1-y_k^{1000})}, \tag{21}$$

$$\sum_{d\in D_4} R_{S_d(4,10)}(x) = \frac{(1-y_k)(y_k^{23}+y_k^{26})(1-y_k^{40})}{(1-x)^2(1-y_k^{10})(1-y_k^{100})}, \tag{22}$$

$$\sum_{d\in D_5} R_{S_d(8,10)}(x) = \frac{(1-y_k)(y_k^{17}+y_k^{19})(1-y_k^{80})}{(1-x)^2(1-y_k^{10})(1-y_k^{100})}, \text{ and} \tag{23}$$

$$\sum_{d\in D_6} R_{S_d(1,10)}(x) = \frac{(1-y_k)(y_k^{83}+y_k^{86})}{(1-x)^2(1-y_k^{100})}. \tag{24}$$

The double sum in (10) is given by (11) and the sum (12) through (24) over all $k \in \mathbb{N}$. Substituting these into (10) gives the desired result. □

It is left as an exercise to the reader to use Theorem 10 to verify that

$$F(x) = 64 + 100x + 136x^2 + 173x^3 + 213x^4 + 253x^5 + 292x^6 + 331x^7 + \cdots.$$

With the assistance of a computer, more of these coefficients have been compiled in Table 8. The table and Theorem 5 indicate that $M_{10^9} < \tau_8 < M_{10^{10}}$ since $\mathcal{W}(M_{10^9}) < \tau_7 < \mathcal{W}(M_{10^{10}})$.

We can now present the value of τ_8. Again, rather anticlimactically, we have

Theorem 11. *The least nonnegative integer of height 8 through $\mathcal{W}$ is*

$$\tau_8 = 101\,\underbrace{373\,373\ldots\ldots\ldots\ldots\ldots373}_{\textit{repeated } 1\,018\,436\,987 \textit{ times}}.$$

n	$\mathcal{W}(M_n)$
10	445
10^2	4674
10^3	53 956
10^4	602 941
10^5	6 890 772

n	$\mathcal{W}(M_n)$
10^6	76 769 074
10^7	831 735 076
10^8	9 179 395 077
10^9	99 655 995 079
10^{10}	1 060 604 995 081

Table 8: A selection of coefficients of $F(x)$ given in Theorem 10.

Proof. We verify that $N = 101 \cdot 1000^{\varepsilon_0+2} + M_{\varepsilon_0}$, where $\varepsilon_0 = 1\ 018\ 436\ 985$, is of height 8. The generating function in Theorem 10 is used to verify that

$$\mathcal{W}(M_{\varepsilon_0}) < \tau_7 < \mathcal{W}(M_{\varepsilon_0+1}) = 101\ 323\ 373\ 385 = \tau_7 + 12. \tag{25}$$

Since $\mathcal{W}(373) = \mathcal{W}(101) + 12$ (via Table 2), then $\mathcal{W}(N) = \tau_7$ and so $\tau_8 \leq N$.

Suppose $n < N$ with $\mathcal{W}^{\#}(n) = 8$. From (25), we know that n must consist of $\varepsilon_0 + 3$ base-1000 digits. Suppose $n = (\alpha_{\varepsilon_0+3}, \ldots, \alpha_1)_{1000}$ with each $\alpha_j > 0$ for $1 \leq j \leq \varepsilon_0 + 3$. If $\alpha_{\varepsilon_0+3} < 101$, then

$$\mathcal{W}(n) = \mathcal{W}(\alpha_{\varepsilon_0+3}) - \mathcal{W}(373) + \mathcal{W}(M_{\varepsilon_0+1}) \leq 13 - 27 + \tau_7 + 12 = \tau_7 - 2.$$

This contradicts the assumption that $\mathcal{W}^{\#}(n) = 8$. Thus $\alpha_{\varepsilon_0+3} = 101$. Taking $\alpha_j < 373$ for any $1 \leq j < \varepsilon_0 + 3$ also leads to a contradiction. And so $n = N$ and the result follows. □

11 Concluding remarks

A summary of the answers is given in Table 9.

There are a number of additional problems that can be explored. The first would be to find additional terms in the sequence (that is, $\tau_9, \tau_{10}, \ldots$). Another is to modify the base function to not count spaces in the word, or to include the word "`and`" where appropriate, or both (as was done in A227290 in Sloane [39]). And finally, another continuation problem would be to ask the same questions in other languages using the counting system prevalent in the associated country based on language and choice of appropriate scale (per Table 3).

12 Acknowledgement

The author is grateful to Google for digitizing many of the older texts [6, 21, 23, 28, 29, 36, 37, 38, 41] that appear in the references.

k	τ_k
0	4
1	0
2	3
3	1
4	11
5	23
6	123
7	101 323 373 373
8	$101\ \underbrace{373\ 373 \ldots\ldots\ldots\ldots 373}_{\text{repeated } 1\,018\,436\,987 \text{ times}}$
9	?

Table 9: Minimal nonnegative integer requiring exactly k iterations of $\mathcal{W}$ to reach the fixed point.

References

[1] W. Abbot, ed., *The Papers of George Washington, Confederation Series*, University Press of Virginia, Charlottesville, VA, 1997.

[2] M. Beeler, R. Gosper, and R. Schroeppel, Memo AIM-239, MIT Artificial Intelligence Laboratory, Cambridge, MA, 1972, http://w3.pppl.gov/~hammett/work/2009/AIM-239-ocr.pdf.

[3] D. Borgmann, *Language on Vacation: An Olio of Orthographical Oddities*, Charles Scribner's Sons, 1965.

[4] D. Borgmann, Naming the numbers, *Word Ways: J. Rec. Linguist.* **1** (1968), 28–31.

[5] D. Borgmann, Renaming the numbers, *Word Ways: J. Rec. Linguist.* **1** (1968), 89–93.

[6] E. Brooks, *The Philosophy of Arithmetic as Developed from the Three Fundamental Processes of Synthesis, Analysis, and Comparison: Containing Also a History of Arithmetic*, Sower, Potts and Company, Philadelphia, PA, 1876.

[7] T. Cai and X. Zhou, On the heights of happy numbers, *Rocky Mountain J. Math.* **38** (2008), 1921–1926.

[8] F. Cajori, *The Teaching and History of Mathematics in the United States*, Washington Government Printing Office, 1890.

[9] J. Candelaria, Extending the number names, *Word Ways: J. Rec. Linguist.* **8** (1975), 141–142.

[10] J. Candelaria, Renaming the extended numbers, *Word Ways: J. Rec. Linguist.* **9** (1976), 39.

[11] J. Conway and R. Guy, *The Book of Numbers*, Springer-Verlag, 1995.

[12] M. Ecker, Strange attractors and black holes (part 5), *REC Newsl.*, Vol. 1, No. 6 (1986), 3–4.

[13] M. Ecker, Mathemagical black holes (part 9), *REC Newsl.*, Vol. 2, No. 6 (1987), 6.

[14] M. Ecker, Number play, calculators, and card tricks: Mathemagical black holes, in E. Berlekamp and T. Rodgers, eds., *The Mathemagician and Pied Puzzler: A Collection in Tribute to Martin Gardner*, A. K. Peters, 1999, pp. 41–52.

[15] G. Flegg, C. Hay, and B. Moss, eds., *Nicolas Chuquet, Renaissance Mathematician*, Reidel Publishing, Dordrecht, Holland, 1985.

[16] M. Gardner, Mathematical games, *Sci. Amer.* **214** (1966), 112–115.

[17] M. Gardner, *The Magic Numbers of Dr. Matrix*, Prometheus, Buffalo, NY, 1985.

[18] I. Greenwood, *Arithmetick, Vulgar and Decimal, with the Application Thereof to a Variety of Cases in Trade and Commerce*, Kneeland and Green, Boston, MA, 1729.

[19] H. Grundman and E. Teeple, Heights of happy numbers and cubic happy numbers, *Fibonacci Quart.* **41** (2003), 301–306.

[20] R. Guy, *Unsolved Problems in Number Theory*, 2nd edition, Springer-Verlag, 1994.

[21] N. Heath, *Treatise on Arithmetic: Through Which the Entire Science Can be Most Expeditiously and Perfectly Learned Without the Aid of a Teacher*, T. Ellwood Chapman, Philadelphia, PA, 1856.

[22] W. Henkle, Names of the periods in numeration, *Ohio Educ. Monthly: J. Sch. Home Educ.* **9** (1860), 151–153.

[23] A. Holbrook, *The Normal: or, Methods of Teaching the Common Branches, Orthoepy, Orthography, Grammar, Geography, Arithmetic and Elocution*, Barnes and Burr, New York, NY, 1859.

[24] E. Kasner and J. Newman, *Mathematics and the Imagination*, Simon and Schuster, NY, 1940.

[25] B. Kolman, R. Busby, and S. Ross, *Discrete Mathematical Structures*, 5th edition, Pearson/Prentice Hall, 2004.

[26] S. Kravitz, The lucky languages, *J. Recreational Math.* **7** (1974), 225–228.

[27] D. Leonard, Harvard's first science professor: A sketch of Isaac Greenwood's life and work, *Harv. Libr. Bull.* **29** (1981), 135–168.

[28] J. Locke, *An Essay Concerning Human Understanding*, Thomas Bassett, London, UK, 1690.

[29] S. Loomis, *Normal Arithmetic: A Text-book, Theoretical and Practical, in Six Parts*, J.B. Lippincott & Co, Philadelphia, PA, 1859.

[30] A. Marre, ed., Notice sur Nicolas Chuquet et son Triparty en la sciences des nombres, *Boll. Bibliogr. Stor. Sci.* **13** (1880), 555–592.

[31] A. Marre, ed., Le Triparty en la sciences des nombres par Maistre Nicolas Chuquet, Parisien, *Boll. Bibliogr. Stor. Sci.* **13** (1880), 593–658, 693–814.

[32] A. Marre, ed., Appendice au Triparty en la sciences des nombres de Nicolas Chuquet, Parisien, *Boll. Bibliogr. Stor. Sci.* **14** (1881), 413–460.

[33] B. Moss, Chuquet's mathematical executor: Could Estinenne de la Roche have changed the history of algebra?, in C. Hay, ed., *Mathematics from Manuscript to Print 1300–1600*, Clarendon Press, Oxford, 1988, pp. 117–126.

[34] L. Noll, How High Can You Count?, http://www.isthe.com/chongo/tech/math/number/howhigh.html.

[35] R. Ondrejka, From unillillion to ultimillillion, *Word Ways: J. Rec. Linguist.* **9** (1976), 177–178.

[36] N. Pike, *A New and Complete System of Arithmetic Composed for the Use Of The Citizens of the United States*, John Mycall, Newburyport, MA, 1788.

[37] N. Pike and C. Dewey, *A New and Complete System of Arithmetick Composed for the Use Of The Citizens of the United States*, 4th edition, Parker, Troy, NY, 1822.

[38] E. Slack, The origin of arithmetical symbols, *Ohio J. Educ.* **2** (1853), 214–221.

[39] N. J. A. Sloane, The On-Line Encyclopedia of Integer Sequences, http://oeis.org.

[40] D. Smith, *History of Mathematics, Vol. 2, Special Topics of Elementary Mathematics*, 1925. Reprinted by Dover, NY, 1958.

[41] R. Southey, *The Doctor, &c, Vol. II*, Longman, Rees, Orme, Brown, Green and Longman, London, UK, 1834.

[42] *Webster's Third New International Dictionary of the English Language Unabridged*, Merriam-Webster, Inc., 1993.

[43] M. Zerger, Fatal attraction, *Math. Comput. Educ.* **27** (1993), 116–123.

2010 *Mathematics Subject Classification*: Primary 00A08; Secondary 05A15, 01A99.
Keywords: recreational mathematics, height, numeration, generating function.

(Concerned with sequences A001273, A005589, A227290, and A301408.)

Received July 6 2018; revised versions received October 1 2018; October 3 2018. Published in *Journal of Integer Sequences*, December 5 2018.

Return to Journal of Integer Sequences home page.