The Beale Cipher

Many students venture into the field of math hoping to one day use their numerical skills to strike it rich, quick and easy. Unfortunately, spending one’s life attempting to decipher the Beale texts will most likely not yield quick and easy riches. The ever trustworthy internet boasts a large number of cryptanalysts claiming to have solved all of the Beale texts but with an equally large number of excuses as to why they have yet to uncover any riches. Scholarly opinion has at best stated a code exists, but the treasure it details is nothing more than a hoax.

The tale of Thomas Jefferson Beale begins in 1817 with Beale and 30 other adventurers heading west to hunt buffalo and grizzly in the vast new frontier. As luck would have it, the very next year Thomas Beale stumbled upon a large store of gold and silver roughly 300 miles north of Santa Fe.¹ Keep in mind this was nearly 30 years before the California Gold Rush.

The men, agreeing to split the treasure equally, worked together to clandestinely move the treasure back to Virginia. They hauled it by wagon and buried it in a rock vault underground approximately four miles from Buford’s Tavern in Bedford County. In 1822, Beale and his cohorts gave a Robert Morriss, an inn-keeper in Lynchburg, an iron box holding the three enciphered texts.² His job was to keep the box safe and distribute the treasure to the relatives of the men in case a tragedy befell the members of the group. Apparently a tragedy did befall the men because by all accounts they are not mentioned any further in this tale. Morriss opened the box 23 years later in 1845.³ When the box was entrusted to him, Beale said a friend from St. Louis would mail him the key to the ciphers. Obviously, the key never made it to Morriss. He

¹ (Nickell 1982)
² (Nickell 1982)
³ (Nickell 1982)
attempted unsuccessfully to solve the ciphers and turned over the ciphers to James B. Ward in 1861.

It was James Ward who successfully solved the second cipher which detailed the contents of the treasure. The second cipher plaintext reads:

>I have deposited in the county of Bedford, about four miles from Buford’s, in an excavation or vault, six feet below the surface of the ground, the following articles, belonging jointly to the parties whose names are given in number three, herewith:

The first deposit consisted of ten hundred and fourteen pounds of gold, and thirty-eight hundred and twelve pounds of silver, deposited Nov. eighteen nineteen. The second was made Dec. eighteen twenty-one, and consisted of nineteen hundred and seven pounds of gold, and twelve hundred and eighty-eight of silver; also jewels, obtained in St. Louis in exchange for silver to save transportation, and valued at thirteen thousand dollars.

The above is securely packed in iron pots, with iron covers. The vault is roughly lined with stone, and the vessels rest on solid stone, and are covered with others. Paper number one describes the exact locality of the vault, so that no difficulty will be had in finding it.

The cipher was encoded using the Declaration of Independence. Cipher number two, as well as one and three, are strings of random numbers. For cipher number 2, the numbers in the cipher text correspond to that number word in the Declaration of Independence. The first letter of that word is the plaintext letter. For example, cipher number 2 starts 115, 73, 24. The 115\textsuperscript{th} word is instituted, I is the plaintext letter. The 73\textsuperscript{rd} word is hold, thus the second plaintext letter is an h. The 24\textsuperscript{th} is another, thus plaintext letter a and so on.
Upon failing to solve the remaining two ciphers, and exhausting much of his time, money and energy, Ward finally made the Beale Ciphers Public by publishing them in a pamphlet in 1885. Since then many a man, from treasure hunter to computer scientist have failed to crack the Beale Ciphers.

At this point we must point out the flaws in our tall tale. The first of which being that Thomas Jefferson Beale most likely did not exist at all. There is no record of him in any census of the Lynchburg area during that time period. It is thought that the name Thomas Jefferson Beale comes from Thomas Jefferson, being the author of the Declaration of Independence, and Beale was the name of the man who brought back the word of gold being found in California. The only person claiming to have ever met him him was Robert Morriss. When Mr. Morriss spoke of Beale to Mr. Ward, he claims that Beale stayed in his hotel in 1820 and 1822. Public Record states that Morriss became the owner of the Washington Inn in 1823, thus there is no way that Beale could have stayed in the inn at the time Ward detailed in his pamphlet.

One theory on the origins of this tale are that the ciphers were made by James Ward himself. This theory is corroborated by stylometric analyses. Stylometric analysis is based on the theory that individuals write in the same style with only slight variations. In the analysis, Thomas Beale’s letters to Morriss were cross referenced with James Wards pamphlet. Also included were the writings of three individuals living in Lynchburg around in the early 1820’s. Thomas Beale was considered to be an educated man and thus the writings of Chief Justice John

---

4 (Nickell 1982)
5 (Nickell 1982)
6 (Nickell 1982)
7 (Nickell 1982)
8 (Nickell 1982)
Marshall and John Randolph (both men of higher learning), and John Randolph Tucker, another man from Lynchburg were used for comparison as well.

The first data sample was the average number of words per sentence. Beale had an average of 29.95, Ward’s was 29.74, Marshall’s was 19.13, Randolph’s was 20.11, and Tucker’s was 34.49. The data shows Beale and Ward were nearly equal in this field while Marshall, Randolph, and Tucker varied a large amount.

The next comparison was the percentage of total words that were the, of, and and. Beale measured 5.1%, 3.51%, and 3.66% respectively. Ward measured 5.2%, 3.26%, and 3.47% again nearly identical. For the record Marshall measured 3.08%, 1.92%, and 2.69%. Randolph measured 4.55%, 2.93%, and 2.87%. Tucker measured 7.35%, 5.28% and, 3.85%. Again, Beale and Ward were very similar in this data sample.

The next form of analysis was the percentage of occurrence of beginning a sentence with the, of, and and. Beale began sentences with the 6.8% of the time. He never began a sentence with of or and. Ward began sentences with the 6.12% of the time. He also never began a sentence with and or of. Marshall began with the 8.93% of the time and of 2.86%. He never started a sentence with and. Randolph started with the 10.34%, never started with of, and used and 1.82% of the time. Tucker used the at the beginning 2.9% of the time, of 1.73%, and 1.54% of the time. Once again, Beale and Ward displayed similar styles in word usage.

Next, the number of commas and semicolons used per sentence was compared between each author. Beale used 2.6 commas and .06 semicolons per sentence. Ward used 2.4 commas and .06 semi-colons per sentence. Marshall used .32 commas and .02 semi-colons per sentence. Randolph used 1.6 commas and .16 semicolons per sentence. Tucker used 2.96 commas and .18
Beale and Ward had nearly identical statistics in this field, yet varied significantly from Marshall, Randolph, and Tucker.

The data above leads one to believe the hypothesis that Thomas Beale and James Ward are the same individual. This begs the question, why did Ward go to all the trouble to create this vast hoax? One theory is points to Wards ties with the free masons. Records indicate that James Ward joined the Lynchburg Chapter of freemasons in 1863. In the decoded text of cipher number two, it speaks of the treasure in a stone lined vault. A stone lined vault is symbolism in the free mason’s society for a secret path in life, beyond that which most individuals take. Thus one could reason that Ward wrote these ciphers as a means to further the goals of the free masons.

Another interesting comparison can be made between Edgar Allen Poe’s work and the Beale Ciphers. Edgar Allen Poe was an avid cryptologist and was known to include codes into his works, *The Gold Bug* being his most famous, and ironically about a treasure. Many of Poe’s preferred themes, phrases, and words appear in the Beale Papers.

Poe often included phrases in foreign languages in his work; roughly 64 out of 70 tales included some reference to a foreign language, his favorite being French. The Beale papers had two references to the French Language: “His house was the par excellence of the town…” and “he was sans peur et sans reproache.” The phrase “Par excellence” was used in 6 of Poe’s stories.

---

9 (Nickell 1982)
10 (Ward 2002)
11 (Ward 2002)
Indefinite or approximate measures also permeated Poe’s work. About 66 out of 70 of his stories included measures such as “perhaps, six feet in height”, “fifth or sixth month” “about three miles long.” Beale used similar types of indefinite measure in his work. Some examples found are “about six feet in height;” “a week or ten days,” and “two or three servants.”

The similarity between Poe and Beale’s writing style continues in they both were heavy users of descriptive conjunctions. They made references to body parts with special focus placed on the eyes. Beale and Poe regularly started sentences with “it was.” There was also Poe’s intrigue with the number 750,000. Poe had estimated the worth of his family’s estate to be approximately $750,000; however he was excluded from the will and reduced to a state of poverty for the remainder of his life. Coincidentally, the Beale Papers valued the treasure to be worth about $750,000.

Perhaps the most intriguing part of the Poe-Beale connection is the similarities between The Journal of Julius Rodman and Thomas Beale’s story. Thomas Beale and Julius Rodman were very similar men. They were young, active, and adventurous men from Virginia. Their stories were almost precise inversions of one another: Beale heading west to hunt and enjoy the wild but instead found treasure, while Rodman headed west to find treasure and instead found a love of the freedom of the west. Both made these discoveries in April (the first spring of their respective trips) after a long winter.

The confidants of Beale and Rodman are also coincidentally similar men. Where Beale had Robert Morriss, Julius Rodman had a man named Andrew Thornton. Thornton and Morriss were considered by their peers to be trustworthy and virtuous men. They were both instructed to

---

12 (Ward 2002)
13 (Ward 2002)
be responsible for dividing the profits of the expedition among the whole group. Thus enough evidence toward the commonality between Poe and Beale’s story should be cause to not dismiss Poe from the whole Beale dilemma.

Let us now examine the number strings themselves. Beale Cipher one has a length of 520. Its highest value is 2906, while its lowest is 1. It has 177 unique values (values that only appear once). There are 298 distinct values (values that appear at least once). The most frequent number to appear was 18 which appeared 8 times. Beale Cipher two has a length of 763 words. The highest value is 1005. The lowest is 1. It has 43 unique values and 180 distinct values. The most frequent number to appear was 807 which appeared 18 times. Beale Cipher number three had a length of 618. Its highest value is 975. Its lowest value is 1. It has 125 unique values and 264 distinct values. The most frequent number was 96 which appeared 13 times.

One of the most in-depth analyses of the Beale Cipher came from Dr. Carl Hammer and his Univac 1108 machine in the late 60’s.\textsuperscript{14} Hammer used three programs, called CRYPTA, CRYPTS, and CRYPTT. Each one used a variety of statistical tests on the number strings of each of the Beale Cyphers. The CRYPTT program produced little valuable data. Dr. Hammer was able to draw some conclusions from the CRYPTA and CRYPTS program.

The CRYPTA program takes strings of random numbers and analyses it with statistical tools.\textsuperscript{15} It lists one by one the runs up and runs down of the numerical strings, sorts them and computes their difference. The data is reduced mod 26 and a frequency table is produced if the results pass a chi-square significance test. The data is tested for randomness using auto regressive analysis which looks for significant cycles in the raw data.

\textsuperscript{14} (Hammer 1970)
\textsuperscript{15} (Hammer 1970)
CRYPTA found that the Beale Cyphers have a deficiency of runs-up of length one, the expected number of runs up of two and a deficiency of longer run-ups compared to the expected value for random number strings. The sorting of the data indicated that the Beale Ciphers seem to have numbers which are Poisson distributed random numbers as oppose to rectangular distributed numbers. This is indicative of some type of human pattern to the Beale number sequences as oppose to just random numbers. The auto regression analysis found that cipher one had much shorter cycles than cipher 2 and cipher 3. This indicates that Beale was much more careful (whether on purpose or accident) in enciphering the first paper. The final piece of evidence that indicates the Beale ciphers were non-random was the mod 26 reduction. This showed that cipher one had a large number of even numbers while cipher 3 had many numbers divisible by 5.

The CRYPTT performed a number of list processing tasks on a given alphabetical key pseudotext under control of a given numerical cypher.\(^{16}\) (Each pseudotext was developed in a different manner in an attempt to replicate a Beale text.) It outputs the number of words and literals on the text data and a count of the elements in the numerical cypher. The alphabetical key text is printed out with word counts indicated over the first letter of every fifth word, then printed with the first letter of every tenth word. Then a frequency analyses of all the digrams (a pattern of two letters next to each other) is printed on a 26 by 26 matrix, AA through ZZ. Next a letter frequency analysis of the keytext is given. The last piece of data outputted is list of numerical cypher elements in an array. The program then takes all this data and using five

\(^{16}\) (Hammer 1970)
different techniques, finds word and letter concordances from the alphabetical pseudotext and the numerical cipher.\textsuperscript{17}

There were three major conclusions Dr. Hammer drew from his statistical testing of the Beale Cipher. First and most importantly, the first and third Beale Ciphers are real codes. They are not just random number strings with the purpose of misleading treasure hunters, but rather do contain some type of information. The method used to encipher one and three is similar to that of cipher two, however letter counting rather than word counting may have been used (spaces and punctuation were not counted). Finally, the Beale Ciphers are written in English.

Another computer test was conducted on Beale Cipher number one by James J. Gillogly. He created a simple computer program that inserted corresponding first letters of the Declaration of Independence to create plain text, similar to James Ward’s method of solving Beale Cipher number two. Recall that Beale number one was composed of 495 values from one to 2906. The Declaration of Independence has only 1322 words in it. To account for this at any point in the cipher a number was greater than 1322, a question mark was inserted. The results of this produced a letter sequence of ABFDE FGHII JKLMM NOHPP.\textsuperscript{18} The first F and the second H could easily be encryption errors. The first F corresponds to 195 of the cipher text, 194 in the declaration is a C. Similarly, the second H corresponds to 301 in the declaration, 302 is the letter O. Dr. Hammer, based on his results from the previous tests found 23 possible errors in the cipher text, thus it is believable that two errors were made in creating the Beale Cipher Number 2.\textsuperscript{19} Regardless, we can simply examine the 14 letter string from D to O. Gillogly explains how unlikely this is for supposedly random text with a simple model. With random text, each letter

\textsuperscript{17} (Hammer 1970)
\textsuperscript{18} (Gillogly 1980)
\textsuperscript{19} (Hammer 1970)
should occur with equal frequency. Thus the probability that the next letter is the same (because of the repeats in the above string) or the next letter in the alphabet is \(2/26=1/13\). From letter two to letter three, the probability is the same, 1/13. Thus the probability for the 3 letter string to repeat or follow the alphabet is \(1/13^2\). The probability of this occurring for a 14 letter string is \(1/13^{13}=3.3\times10^{-15}\). By accounting for the 495 values we get \(495/13^{13}=1.6\times10^{-12}\), in other words, highly unlikely. Gillogly concluded from this that the Beale Cipher number one was a hoax.\(^{20}\) Whoever made it used the Declaration of Independence to create random text, and decided to throw alphabetical strings in due to boredom.

As we are now no closer to solving the Beale cipher then we were at the beginning of this paper, we can at least derive a few theories behind the probable hoax. There is no Thomas Beale, however there are strong cases to be made that James Ward or Edgar Allan Poe played a role in producing the story. The existence of a treasure is unlikely at best, but the existence of a cipher in papers one and three is not unreasonable. Cryptologist and treasure hunter alike will continue to toil on the Beale mystery and perhaps will one day discover the secret behind the numbers.

\(^{20}\) (Gillogly 1980)
Works Cited


