5

SIGNATURE SIMULATION

AND CERTAIN CRYPTOGRAPHIC CODES

Carl Hammer, Ph.D.
Director

Computer Sciences
UNIVAC
Washington, D.C.

[illegible]

Invited Paper

Third Annual Simulation Symposium
14 January 1970
Tampa, Florida

Signature Simulation and Certain Cryptographic Codes

Carl Hammer, Ph.D.
Director, Computer Sciences
UNIVAC, Washington, D.C.

Abstract

Three cyphers allegedly authored by Thomas Jefferson Beale in 1822 have been the subject of intensive study for over one hundred years. Generations of cryptanalysts have expended untold man-years, thus far without success, attempting to decode them; vast armies of fortune hunters and treasure seekers have devoted Herculean labors to digging up the rolling hills of Virginia trying to locate the promised bonanza. The history of pertinent activities would fill volumes yet serious students of cryptography have always had nagging doubts about the cyphers' authenticity. It has been alleged that the "known solution" to Cypher Number Two: 115, 73, 24, 818, 37, 52, 49, ... ("I have deposited in the County of Bedford about four miles from Buford's in an excavation or vault...") with the aid of an unsanitized version of the Declaration of Independence was merely a superb, imaginative and grandiose hoax perpetrated ages ago for whatever reasons.

Modern computer technology could obviously perform signature analyses on the Beale cyphers and could also, in fact, simulate the process of encoding itself so as to yield new clues and deeper insights into their construction. For the benefit of the uninitiated, the encoding method used in the second cypher employs a specified document whose words are simply numbered consecutively and first letters of these words are sought out at random to match the letters of the cleartext or message. The sequence of numbers corresponding to these matches is then written down as the final code. While primitive, the process has the advantage of relative security until the source document becomes known; at that moment the cypher can be decoded even by second graders.

The work now completed with the help of our UNIVAC 1108 includes numerous analytical studies of the Beale cyphers and various types of simulations. For example, we have turned the entire process of simulated encoding by various schemes over to the machine and analyzed the signatures of these synthetic codes; we have also encoded various messages by hand, using different texts and a variety of methods to obtain their signatures. These simulations provide convincing evidence that the signatures are both process and data dependent; they indicate also very strongly that Mr. Beale's cyphers are-for real and that it is merely a matter of time before someone finds the correct source document and locates the right vault in the Commonwealth of Virginia.

Table of Contents

1. Introduction

These cyphers allegedly authored by one Thomas Jefferson Beale in 1822 have been subject of intensive study for over one hundred years. Generations of cryptanalysts have expended untold man-years attempting to decode them while treasure hunters have spent an equal amount of time and effort in digging through the hills and caves of Virginia in an attempt to locate Beale's treasure. During the summer of 1968, several members of the American Cryptogram Association (ACA) decided that a concentrated group study might be successful where individual efforts had thus far failed. In response to several inquries, eleven persons indicating an interest in this cypher convened in Washington on Saturday, 20 September 1968, to discuss present knowledge, pool talents and resources, and formulate plans for future work. It was unanimously agreed that modern computers should be used to analyze the content of these cyphers in depth, to develop their "signatures," and to simulate the encoding process allegedly used by Beale in his three messages. The group suggested numerous modifications of already existing analytical computer programs and the ideas proposed then were eventually translated into real and working programs.

This report summarizes the work done since then. Naturally, we cannot include all the detailed computer printouts which have accumulated. These printouts, however, can be made available for inspection in our Washington office at any time. We hope that interested parties will avail themselves of this opportunity and that during the year 1970 joint efforts will bring us closer to the solution of this very interesting project.

As previously mentioned, the Beale Cyphers are three numerical codes allegedly constructed during the second decade of the past century by Thomas Jefferson Beale for the purpose of identifying the site of a treasure buried by him. The three codes are shown in Appendix 9. The method used by Beale to encode the second of his three messages was "broken" by a James B. Ward several decades later. It is very much like that already described by Sir Arthur Conan Doyle in "The Valley of Fear." Taking any readily available source document, such as the Declaration of Independence, each word in this keytext is numbered sequentially: (1) When (2) in (3) the (4) course (5) of (6) human (7) events ... The letters of the message to be encoded are then selected at random from appropriate starting letters of these words. The final code consists thus only of a string of numbers, as in Beale Cypher Number 2: 115, 73, 24, 818, 37, 52, ... Correlating these numbers against the keytext, we find that they represent consecutively the letters I, H. A, V, E, ... and this combination of keytext and code reveals quite readily the entire message contained in B2:

> 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 November 1819. The second was made December 1821 and consisted of nineteen hundred and seven pounds of gold and twelve hundred and eighty eight pounds of silver, also jewels obtained

3. Simulation Studies and Synthetic Codes

One fundamental question which has permeated the history of the Beale Cypher has been a determination of its authenticity. There are some who believe that it is nothing more than a grandiose hoax, while others firmly believe it is legitimate and will be cracked sooner or later. In order to arrive at an answer to this question, one might proceed from the assumption (statistical hypothesis) that Beale Cyphers 1 and 3 are random doodles while Beale Cypher 2 was constructed to create the basis for a hoax. If it now could be proven that Cyphers 1 and 3 were purely random numbers with a "signature" significantly different from that of Cypher 2, the weight of the evidence would tip the scales toward abandoning hope of ever obtaining legitimate solutions to Cyphers 1 and 3. Several studies along these lines have been conducted and are discussed below.

3.1 Code Simulation with Rectangular Random Numbers

The most primitive assumption which we could make is that Beale Cyphers 1 and/or 3 were written down as sequences of pure random numbers. For exampl they could be rectangular random numbers with a given range. In order to test such an hypothesis, we wrote a short Fortran Program to produce strings of rectangular random numbers which are then punched into data cards having the same format as those acceptable to the CRYPTA and CRYPTT Programs. By way of defining rectangular random numbers, note that they have a uniform distribution of such a nature that the occurrence of any number within the stated range is equally likely. As a consequence, the sorted array of thes numbers will have no mode and their first differences also tend to be uniformly distributed. The subject program was used to generate one output deck RS (for Rectangular Simulation) which we used to test the earlier stated hypothesis (cf. Section 4.4).

3.2 Code Simulation with Poisson Random Numbers

Even a most cursory inspection of the three Beale Cyphers B1, B2, and B3 reveals that the numbers are not uniformly distributed. Therefore, a second random number generator program was developed whose output is Poisson distributed. This distribution occurs widely in psychological tests, communications, and other natural and engineering phenomena. According to this distribution law, smaller numbers will occur more frequently than larger numbers and one such set of data was generated. The output deck PS (for Poisson Simulation) is arranged in a format acceptable to the CRYPTA and CRYPTT Programs. The results obtained with PS are described in Section 4.5.

3.3 Hammer's Simulation of the Beale Process

The data decks RS and PS described above are strictly statistical.simulations of some assumed random number districution law, permitting us only to test the Beale Cyphers against these hypothetical distributions. It was felt that we should go one step further and actually simulate the Beale process itself.

frequency zero among the first letters of the keytext words, the next letter in the cyclical alphabet, A, B, C, ... Z, A, B, .. is chosen to replace any unencodable letter.

2.3 CRYPTT Program Description

This Fortran program performs a number of list processing tasks on a given alphabetical keytext under control of a given numerical cypher. In its present form two punched card input decks are required. The first deck begins with a text header card from which later program outputs are constructed; it is followed by a keytext data deck and an extra control card which signifies its end. The second data deck begins also with a text header card which is followed by cards containing the numerical cypher data. The last card of the numerical cypher data deck carries a control punch signifying its end.

The program outputs first list the two headers to identify the source of the alphabetical keytext and of the numerical cypher data. They are followed by a summary giving the number of words and literals in the textual data and a count of the elements in the numerical cypher. The alphabetical keytext is then printed out with word counts indicated over the first letter of every fifth word. It is printed out again with a letter count indicated over every tenth letter; spaces are not included in this count. Then follows a digram analysis of the text which tabulates frequencies of the digrams in a 26x26 matrix ranging from AA to ZZ with row and column totals; the digram list is also printed out in order of descending frequencies. A letter frequency analysis of the keytext is supplied. Finally, a listing of the numerical cypher elements is given in array form.

At this point the program develops several letter and word concordances from the alphabetical text and the numerical cypher. The first parameterized approach matches the numerical entries of the cypher against corresponding first letters of the search text; it also introduces integer lags into this matching process and the key cypher numbers are systematically incremented by these lags. The outputs from this first approach are printed as pseudotext where blanks replace impossible word or character numbers. In the second approach, the parameterized matching process creates a pseudotext by reversing the first process and printing the last characters of the words matching the given cypher numbers. Again, an arbitrary lag is introduced both in the word and character counts. The resultant pseudotext is then printed out, allowing for blanks in impossible word or character assignments. In the third approach, the key elements of the cypher are taken to be word position counts. Pseudotext is again created and printed out with an arbitrary incremental lag imposed on the numerical cyphor. In the fourth approach, the program matches letter position numbers with arbitrary lags against the numerical cypher. In this approach blanks in the text are not counted. Finally, in the fifth approach, a match is established between the numerical cypher elements and character counts which include blanks. Typical running times of the program with 8000 literals, 500 numerical cypher key elements, and with lags ranging from 0 to 10 are less than one minute on the Univac 1108 system.

> in St. Louis in exchange 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 stones, and the vessels rest on solid stones 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.

This is not a tutorial on cryptography and we shall discuss here only the methodology employed in this particular encoding/decoding process. First of all, it is obvious that even if the exact methodology were known, decoding without an exact specification of the keytext may be a very difficult process. Even with the help of advanced cryptographic methods it can introduce obstacles of enormous magnitude. Secondly, the encoding method indicated by Sir Arthur Conan Doyle and Beale's B2 is only one of several possible variations. For example, instead of the first letter of the numbered words, their second letters could be chosen, or their last letter, etc. Then, instead of numbering the words of the basic document, the letters could be numbered sequentially counting or not counting blanks and/or punctuations. The encoder may also introduce a lead or lag function λ such that the code element N actually refers to word or letter number $N \pm \lambda$. At this point it becomes clear that we have at least a major data processing problem on our hands when we encounter cyphers of this type. In fact, it is more than likely that we also have a major cryptographic problem if very little is known about the source of the cypher. In the case of historical cyphers, there is the additional difficulty of locating the authentic documents or, what may even be worse, unsanitized or specialized versions thereof that an author of centuries past may have used. More will be said later about this problem in connection with our own work on Beale Cyphers 1 and 3.

2. Three Computer Programs

The power of the Univac 1108 machine was tapped with the aid of several computer programs specially developed for our purpose. The tasks which these programs carry out fall into three categories. The first CRYPTA Program is basically analytical; it takes a string of numbers (i.e, a numerical code) and analyzes it with the help of many methematical-statistical tools. The second CRYPTT Program involves list processing and various decoding attempts at obtaining a concordance between a given numerical code and an alphabetical keytext. The third CRYPTS Program is a computer simulation of the human process of encoding some cleartext with the help of a given alphabetical keytext.

2.1 CRYPTA Program Description

This Fortran program performs a number of analytical tasks on a given numerical cypher. In its present form, the inputs are punched cards, the last of which carries a special punch to signify the end of the data deck. The data deck is preceded by one informational BCD card which is used to construct the heading of the outputs.

The program outputs begin with a summary of the data statistics which includes the title (from the header card), the number of entries in the cypher, their numerical average, their root-mean-square, as well as a listing of the raw data. Next, runs-up and runs-down are enumerated, followed by a sort of the data and their first differences. The original data are then reduced modulo 26 (an attempt to correlate them with the English alphabet) and the resultant frequency table is printed out and tested against English letter frequencies. This test is then repeated for all possible 26 cyclical permutations but results are only printed out if they submit to a Chi-Square statistical significance test. The data are also cross-summed and the resultant array is printed out. Again, the cross-summed frequency table is compared with English letter frequencies and the results are printed out if they are statistically significant.

Next, the data are subjected to an autoregressive analysis which looks for statistically significant cycles "hidden away" in the raw data. Significant frame sizes are printed out for autoregressive lags ranging from 2 to 30. Respective averages and standard deviations are given for each frame position in statistically significant frame sizes. Finally, a Kasiski-type analysis is performed on the raw data elements by examining their differenced position values in the cypher. This analysis is summarized by listing frequencies of the divisors ranging from 2 to 36. Typical running time of this program with 500 data points, including compilation, is about eleven seconds on the Univac 1108.

2.2 CRYPTS Program Description

This Fortran program encodes a given alphabetical cleartext (which may not contain any numbers or special symbols) with the help of another alphabetical keytext by the concordance or matching process allegedly employed in the Beale cypher. Program output is a listing and a deck of punched cards in the same format used for input into the CRYPTA and CRYPTT programs described elsewhere. This program now has three options.

The first option searches the keytext sequentially, always beginning with its first word, until a match between a given cleartext letter and the first letter of some word in the keytext is obtained; the position number of that word is then recorded. Encoding of the cleartext will thus produce a string of the lowest valued position numbers in the keytext. If no match is found during the complete search of the keytext, dummy numbers 1, 2, 3, .. are successively inserted into that string.

The second option searches the keytext sequentially, beginning with its first word, until a match between a given cleartext letter and the first letter of some word in the keytext is found; the position number of that word is then recorded. However, when the same type letter comes up again for encoding the search for a match resumes at the position last recorded and this process is continued to the end of the keytext before returning to its beginning. If a letter to be matched does not occur in the first letters of the keytext words, the next letter in the alphabet is chosen cyclically, i.e., Z is followed by A.

The third option uses a rectangular random number generator to select matching first letters from the keytext words in the process of encoding the cleartext letters. If a letter required in the encoding process has

We chose a paragraph of text (randomly from a speech recently published by this writer) and proceeded to encode it by the alleged Beale Process. A listing of Beale's Version of the Declaration of Independence was used as the keytext for encoding this cleartext. The annotated keytext carried word numbers for every fifth word throughout and was printed on typical computer output paper. As a matter of fact, it was the by-product of one of the earlier CRYPTT runs.

We scanned the text sequentially to find the required letters, writing dowr word position numbers as their first letters were found to match the looked for letters. We would proceed for a while in this fashion then turn to another section of the keytext and continue the sequential search-and-match-process. While carrying out this work, we noted the mental strain in searching for rare or non-existent letters. We were also tempted to "memorize" position numbers for certain letters and we developed a resistance to turning to later sections of the keytext. However, we made a conscious effort to switch to different areas of the text, not previously used, and we worke the text always in a forward-search mode, never in reverse. Naturally, al these psychological factors would be reflected in the "signature" of the produced string of numbers, as indicated by the analysis in Section 4.6.

3.4 Caldwell's Random Data Code

One member of the study team who submitted a "personalized" random data code was Mr. Robert Caldwell. Nothing is known (on purpose) about the method which he used to produce these CS data. Upon receipt of his manuscript, data cards were punched, using the format acceptable to our CRYPTA and CRYPTT Programs. These decks were then subjected to computer analysis.

3.5 Nelson's Random Data Code

Mr. Carl Nelson, another member of the study team, has also submitted a "personalized" random data code NS about whose source we did not inquire. His code, too, was punched up and subjected to analysis by the CRYPTA Program.

3.6 Three Synthetic Codes Generated with CRYPTS

It was a simple matter to take the same text used in Hammer's simulation HS and encode it with the three CRYPTS-options. The three optional outputs S1, S2, and S3 could then be analyzed with the help of CRYPTA (Section 4.9) and could be validated by running them against CRYPTT. Naturally, there would be no startling results forthcoming from this last experiment since the three codes were known to have a solution. They would serve only as a benchmark in the analysis of codes B1 and B3 of unknown origin.

4. Analytical Studies with the CRYPTA Program

Including the three original Beale Cyphers, our stockpile of real or simulated codes now contains eleven sets of data (cf. Section 3) allowing exhaustive analysis with the CRYPTA program. Of maximum interest, of course, is a comparison of measurable statistical parameters *for simulated and* real codes. We shall highlight significant differences in these parameters reflecting structural or signature information in the following.

4.1 Beale Cypher No. 1

The B1 cypher has 520 entries with a mean of 273 ranging from 1 to 2906 (cf. Appendix 9.1) The distribution of runs-up-and-down varies significantly from random data; there is an excess of runs-down of length one which is compensated for by a shortage of runs down of lengths three and higher. This provides a possible clue to the construction of the cypher as the author might have "jumped back" more frequently while engaged in the encoding process. Autoregressive analysis reveals only one significant pattern (of length sixteen) but by itself this fact does not give rise to any suspicions about the nature of the code. Modulo reductions of the code numbers yields twelve significant variations from randomness indicative of the difficulties which the author might have experienced in selecting keytext letters to encode his message. Significantly, most of these parameters are even; this fact might indicate that the author numbered only every other word of his keytext and then fell for the psychological preference for numbered over unnumbered words.

4.2 Beale Cypher No. 2

The B2 cypher is longer than B1. It has 763 entries ranging from 1 to 994 with a mean of 162 (cf. Appendix 9.2). It has, of course, a known solution with Beale's Version of the Declaration of Independence as keytext. Again runs-down of length one dominate and are compensated for by too few runs-down of length three and up. Autoregression reveals three significant cycles of lengths three, five, and seventeen which is about right for a hand-coded job. Modulo reduction indicates again a significant deviation from randomness with 21 parameters but this time only half of them are for even numbers, indicating that the method of encoding chosen was "better balanced."

4.3 Beale Cypher No. 3

The length of cypher B3 falls between B1 and B2. It has 618 entries ranging from 1 to 975 with a mean of 153. However, runs-up-and-down extend far beyond the range to be expected from random numbers with three runs-up of length nine and also a shortage of runs-down of length three and greater. In view of this we might suspect that earlier "practice" or a change in the mental approach to the encoding task could account for this unusual pattern. Also, letter-counting instead of word-counting might produce such a pattern. Autoregression produces four significant cycles of lengths three, five, eleven, and seventeen. By comparison with B1 and B2 we find that the longest of these cycles reflects the personal "signature" of the author. Modulo reductions of the cypher entries yield six significant values without predominance of odd or even; this time multiples of five dominate, possibly indicating that the numbering scheme used for encoding of this cypher differs from that used in B1 or B2.

4.4 Rectangular Random Number Code

This first of the computer-generated benchmarks has 500 entries whose mean is similar to that of B1 (by design). While there is nothing interesting about the runs-up-and-down, the logarithmic sort fit yields parameters significantly different from the three Beale Cyphers, indicating clearly that the latter come from a non-random source. Autoregression yields only two significant frequencies of 2 and 3; these can be traced to the periodic nature of the random number generator itself. Modulo reduction, likewise, yields only one value of significance against many values for the real cyphers. At this point it becomes thus obvious that none of the Beale Cyphers was constructed with the help of early random number tables, or by tossing coins or rolling dice.

4.5 Poisson Random Number Code

The generation of Poisson distributed random numbers yields several surprises. While their runs-up-and-down are similar to those obtained from the uniform random numbers, the logarithmic sort fit resembles much more the data obtained from the three Beale Cyphers. Autoregression yields three significant values of 3, 7, and 23. The individual values can be traced to the idiosyncrasies of our random number generator. The fact that three of them show up yields a good benchmark for comparison with other simulation schemes. Finally, modulo reduction yields no significant values which indicates once more the non-randomness of the Beale Cyphers.

4.6 Hammer's Simulated Beale Type Data

In addition to having a known solution, these data correspond most closely to B3. There is a preponderance of runs-down of length one, compensated for by too few runs-down of lengths greater than three. This is evidence for the manner in which the encoding process was carried out, namely by running forward along the keytext, and jumping back whenever the urge struck us. Evidently, at least this encoder was unable to control his tendencies in that direction, producing long runs-up against short runs-down. Autoregression yields also five significant cycles, including two of great length, as did B3. On the other hand, modulo reduction produced only two significant values, due probably to the fact that we tried very hard to avoid a preference for the numbered elements of the comparison text.

4.7 Caldwell's Random Data Code

The source of these data and the method of their generation is not known. Comparison with several benchmarks of computer-generated or man-made simulated codes would indicate that they correspond most nearly to RS or S3, which suggests that their basis is indeed a set of true random numbers. Runs-up-and-down, autoregressive parameters, and modulo reduction all point in that direction.

4.8 Nelson's Random Data Code

Here we have an entirely different pattern from the one generated by the Caldwell data. While runs-up-and-down (except for one run-down of length 6) resemble most clearly a random pattern (say, of the Poisson type), autogressive analysis reveals the non-randomness of the data to be very much like Hammer's simulation or the original Beale Cyphers. Also, modulo reduction produces seven values of significance, distributed very much like those of B3. All this evidence leads us to suspect that this cypher is based upon a real text and contains a real message.

4.9 Three Codes Generated with CRYPTS

As mentioned in Section 3.6, this program has three options and it simulates (on the computer) the encoding process that a human being might want to employ. The three codes produce no spectacular or unexpected results. Runs-up-and-down for option one (after a hit, return to beginning of keytext) indicate a dominance of runs-down for length one, as expected; frequencies of longer runs adjust themselves accordingly. Only option two (scan entire text and then return to beginning) yields significant cycles for autoregression. The logarithmic sort fit for option three (uniform selection of matching letters) produces constants which differ significantly from those obtained by the other options. Again, option one yields a large number of significant parameters (22) during modulo reduction as compared with just one such parameter for the other two options. These results proved to be very valuable when we tried to "bracket" the unknown cyphers B1, B3, CS, and NS (cf..Section 6).

5. Decoding Studies with CRYPTT

As mentioned earlier, this computer program eliminates the rote and drudgery connected with setting up concordances between a given numerical cypher (such as the Beale Codes) and a keytext which might yield the cleartext. The CRYPTT Program provides additional statistics about the chosen keytext, of value in further analytical studies. The CRYPTT output subroutines yield only interesting garbage unless the chosen keytext happens to be the "correct" one. However, even the "signature" of this garbage can be rather revealing as we discovered.

5.1 Beale Cyphers Nos. 1 through 3

Beale's Version of the Declaration of Independence (See Appendix 9.4) as keytext produces alphabetical strings but little intelligence for any of the available CRYPTT methods or their variants. For example, if we correlate the numbers of the cypher with the initial letters of correspondingly numbered words in the keytext, we obtain SCS ETFA GSDOTTUCWOTWTAAIWDBIIDTT WTTAABBPLAAABWCT... The spaces result from code numbers exceeding the number of words in the keytext; the computer program maintains a count of these words and inserts blanks where code numbers exceed the upper limit of numbered text words. Letter frequency counts indicate that this pseudotext does not agree with the letter frequency counts of ordinary English. Similarly, if we number the letters in the basic text, rather than the words, CRYPTT produces another string of seemingly random letters:

EONECTONBOESOOHCDELSCSTBOSSHWHEERSCLCSDEESTAILMCAREHD... There may occur chance digraphs and trigraphs in such a string of letters; the partial "pseudotext" shown above contains such words as one, ton, so, boss, tail, care, etc. The letter frequency distribution in this pseudotext matches that of the English language much more closely; also, there are no blanks since the keytext has 6527 literals (not counting spaces) and the highest code number in Beale Cypher No. 1 is 2906.

No further progress is made by applying lags or leads to this cypher. For example, its beginning of: 71, 194, 38, 1701, ... could refer to the first letters of words 70, 193, 37, 1700, ... or to letters 73, 196, 40, 1703, ... etc. However, analysis of the pseudotexts obtained with such lags or leads indicates a much better match with English letter frequencies for the method of letter-counting over the method of word-counting. In this connection, remember that the method of word-counting is the one that "breaks" Beale Cypher No. 2. Nevertheless, letter-counting seems to provide a better option for decoding B1.

Beale Cypher No. 2 can be "broken" by applying the method of unlagged word-counts against the keytext of the Declaration of Independence. All other methods (with or without lags and leads) available under CRYPTT produce interesting benchmarks for letter frequencies against which we can test other decoding methods. For example, if another keytext produces letter frequencies that are statistically less significant than those produced by the Declaration of Independence plus an inapplicable method, we may take this clue to mean that we have the wrong method and/or the wrong keytext. More about this interesting facet of our signature analysis in Section 6.

Beale Cypher No. 3 does not yield anything new and startling. Neither word-counting nor letter-counting produces anything but gibberish but the relevant letter frequencies are again better for the letter-counting method.

5.2 Hammer's Simulated Beale Type Data

These data are based upon a cleartext of modern English writing; their submission to CRYPTT with the Declaration of Independence produces, of course, the correct solution. It furnishes us also with benchmarks for the distortion of letter frequencies resulting from introduction of lags or leads or an incorrect encoding method. For example, decoding by the correct method will produce the cleartext: COMPUTERSHAVESTARTEDTODOAMAZINGTHINGSTHEYDESIGNAIRPLANESGUIDEMISSILES... Introduction of a lead of one unit produces the following pseudo-cleartext: AFIESOLEEIAXTRCEOXAOPIPNEMISEOOATHOAXNEVET... As before, code numbers in excess of the available word-count in the keytext are translated by the computer into blanks. Very few recognizable digraphs or trigraphs result from this simple shift; similarly, letter frequencies are immediately distorted from that expected for ordinary English.

5.3 Caldwell's Random Data Code

Application of the Declaration of Independence against these data produces meaningless letter strings for all methods and relevant lags or leads. For the unlagged letter-count method, we have BOMHHLOLNS T B WP PJTTA TDA NDOFWN TAIPIAD T IPG R S ... Both the frequency of recognizable digraphs or trigraphs is very low as is the correlation of resultant letter frequencies against those of the English language alphabet. Our earlier stated suspicion thus receives support: In all likelihood these were-true random numbers. They are certainly not derived from a cleartext encoded against the Declaration of Independence by the Beale method.

5.4 Nelson's Random Data Code

The Nelson Cypher, unlike Caldwell's Random Data Code, produces strings of letters with frequencies similar to those obtained in other attempts where we deliberately matched data against the "wrong" keytext. For example, the word-counting method, without lag or lead, yields against the Declaration of Independence: OTTSO SOWOC RTRNHSAWAOTI THTBAATLAONIPAANR... Letter frequency counts of the pseudotexts tend to confirm that we have here a legitimate cypher but the wrong document. More about that under Section 6.

5.5 Codes Generated with CRYPTS

Application of a computer-simulated method of encoding will, of course, yield the desired cleartext for the right choice of method. Thus the several options of CRYPTS will lead us to the correct solution, say for the unlagged word-counting method: COMPUTERSHAVESTARTEDTODOAMAZINGTHINGS... which converts for a lag of one into: OFAENHVEEUNAVEHNEHVIHFIF... Letter frequencies here indicate clearly the transition from solution to non-solution even though we have the right text. Again, this result will prove of interest when we attempt to summarize our results in Section 6.

5.6 The Magna Carta

During the latter part of 1965 we had a visitor who stated in no uncertain terms that he had "broken" Beale Codes 1 and 3 but needed the assistance of our computer to complete his work. We ignored the "minor" contradiction contained in his statement and pressed him for further information.

He told us, reluctantly, to "try the Magna Carta" -- and disappeared. We syllogized the following conclusions: (i) He was telling us the simple truth; it is the Magna Carta, (ii) He was leading us down the garden path; it is not the Magna Carta, (iii) He was being super-devious; it is the Magna Carta but he thinks that we think he gave us a false lead, thus we won't try it.

At least this document would provide us with another simulation tool which we had not tried before. In what language was the source document written? If we could make any positive statements by additional, statistical analyses, we would indeed have another piece of information needed to solve the Beale mystery. We therefore studied in depth decoding attempts with CRYPTT, using several versions of the Magna Carta as possible source documents.

There is no need to relate in detail the extensive literature available on the Magna Carta. There exist many versions authored by John in AD 1215 and by Henry in AD 1225. Both documents are in Latin'and both begin with a lengthy Preamble which constitutes.the King's authority and lists his fiefs and advisors. Authorized translations were used by Thomas Jefferson and his friends in the drafting of our own Constitution. If the Beale legend is not myth, the author of these cyphers would have known these documents. A short, sanitized version (both in Latin and English) of John's Preamble can also be found in the Encyclopaedia Britannica.

Table 1 lists some details about the four Latin and the eight English versions of the Magna Carta which were used as keytexts in relevant decoding attempts of Beale Cyphers 1 and 3, discussed in Section 6.2. For the sake of completeness, we have also listed two additional documents, the Declaration of Independence and a modern English text. These were used to establish further benchmarks and to allow us a better assessment of the results of the other CRYPTT runs. The indicated run numbers in the first column of that table refer respectively to runs made against Beale Cyphers 1, 2, or 3, as shown in the second column of the table. Some data about these keytexts, such as number of words or number of letters, are provided in later columns, The table also indicates the source documents and any variants chosen for a particular run. The results obtained from these and additional runs will be discussed in Section 6.2.

6. Analysis of Results

Our studies are largely predicated upon the two programs CRYPTA and CRYPTT and upon computer runs made with the several data decks described earlier. Having already mentioned briefly some of our findings, we shall now try to consolidate our position and review pertinent data in detail.

6.1 Summary of CRYPTA Data

Table 2 provides all of the data obtained for the three Beale Cyphers (B1, B2, B3), Rectangular and Poisson Distributed Random Data Simulations (RS, PS), Simulations contributed by Hammer, Caldwell, and Nelson (HS, CS, NS), and Computer Generated Simulations obtained from the three CRYPTS Program Options (S1, S2, S3).

The first segment of Table 2 gives the number N of data elements in the respective codes, their mean, standard deviation, and range. The second segment of the table summarizes the number or runs-up-and-down. The third segment indicates parameters obtained for the logarithmic sort fit: Intercept A, slope B, and standard deviation S. The fourth segment lists significant cycles detected for autoregressive analysis by length and number; thus 3,5,17(3) for Beale Cypher B2 indicates 3 significant cycles of lengths three, five and seventeen. The fifth segment of the table indicates

Runs	Beale	Language	Source Document	Author	Date	Document Source	Words	Letters	Overall Description, Comments
1-21-41	1-2-3	Latin	Magna Carta	John	12	Encyclopaedia	180	1253	Preamble Only
2-22-42	1-2-3	Latin	Magna Carta	John	121	Stubbs, p. 292	363	2337	Preamble and Text
3-23-43	1-2-3	Latin	Magna Carta	Henry	1225	Stubbs, p. 350	284	1725	Preamble and Text
4-24-44	1-2-3	Latin	Magna Carta	John	1215	Stubbs, p. 284	344	1990	Preamble Omitted
5-25-45	1-2-3	English	Magna Carta	John	1215	Encyclopaedia	228	1140	Preamble Only
6-26-46	1-2-3	English	Magna Carta	Henry	1225	Swindler	676	2852	Preamble and Text; Itali no brackets
7-27-47	1-2-3	English	Magna Carta	John	1215	Swindler	692	3010	Preamble and Text; Brack no italics
8-28-48	1-2-3	English	Magna Carta	Henry	1225	Swindler	671	2806	Preamble Omitted; Bold, italics
9-29-49	1-2-3	English	Magna Carta	Henry	1225	Swindler	456	2280	Preamble Omitted; no brackets
0-30-50	1-2-3	English	Magna Carta	John	1215	Swindler	576	2390	Preamble Omitted; Bold, no italics
1-31-51	1-2-3	English	Magna Carta	John	1215	Swindler	355	1439	Preamble Omitted; Bold, brackets, italics.
12-32-52	1-2-3	English	Magna Carta	John	1215	Swindler	269	1100	Preamble Omitted; III Or "Forever" one word; bra italics; variation on r 11-31-51
13-33-53	1-2-3	English	Declaration	Beale	1789	Committee	1322	6527	Beale's version; Declar of Independence
4-34-54	1-2-3	English	Essay	Hammer	1968	UNIVAC	522	2656	Modern Text; Calibratio

results of the Kasiski analysis; it lists excesses for significant divisors. For example, under Beale Cypher 2 we find only one such divisor, namely 7, which occurs to a greater degree than can be expected if the data base were random. The sixth and final segment of the table summarizes results obtained from modulo reduction; significant moduli are listed and counted. For example, Beale Cypher 3 has six significant moduli, namely 5, 10, 15, 20, 21 and 25, which occur too frequently and must be attributed to some property of the data.

A grandstand view of this table allows us to compare codes with similar properties and to classify these codes as falling between other codes with known properties. If we look at runs-up-and-down, we observe that all Beale Cyphers have a significant deficiency of runs-up of length one, the expected number of runs of length two, and a significant deficiency of longer runs-down. Now we note that a similar pattern obtains for HS and S1, indicating that in constructing these codes Beale and Hammer must have acted very much alike. Recalling our earlier comments (Section 4.6) about the psychological forces which become active when endoding messages by this method, we suspect that Beale must have worked, as we did, with a long document but that he jumped backward more often than forward. In other words, he worked his way down the document, then jumped to another spot, worked down again, but probably never reversed his process by working concordances backwards or up.

The logarithmic sort fit indicates that the Beale Cyphers resemble much more closely Poisson-distributed Random Numbers (PS) than rectangular ones. The S2 technique (which scans the entire keytext for concordances before returning to the starting point) falls short of all three Beale Cyphers; the S1 technique (which returns after each hit to the starting point) rests only above them, there being little significant difference between the observed slopes. Again, we have here an indication that Beale proceeded in jumps, like a human, and less efficient than our patient computer.

Autoregression indicates that Mr. Beale was rather successful in Cypher 1 as he chose his matches without producing significant short cycles. However, B2 and B3 do contain hidden cycles which resemble very much those of Hammer's simulation HS and are bracketed by computer simulations S1 and S2, as before. Here we also find that Nelson's data (NS) look very much like those of a real code, while Caldwell (CS) might have used a very efficient computer random number generator or good tables.

Kasiski analyses were only included on the outside chance that Beale Cyphers 1 and 3 had been constructed by an entirely different method than B2. The absence of significantly abundant divisors in B1 and the two divisors of 14 and 19 under B3 may be considered as spurious, just as the several divisors detected in the simulations. This finding rules out a large class of encoding methods commonly employed during that time, including Beaufort, Gronsfeld, Porta, and other periodic cyphers.

Finally, the results of modulo reductions indicate once more rather strongly the non-random nature of the Beale Cyphers. All three exhibit a significant preference for certain numbers. B1 abounds in even numbers, while B3 prefers numbers divisibly by 5. No such clearcut separation exists for B2

TABLE 2

COMPARATIVE STATISTICS: CRYPTA - SIGNATURE ANALYSIS

IDENTIFICATION		B1	B2**	B3	RS	PS	HS**	CS	NS	S1**	S2**	S3**
N		520	763	618	500	500	424	500	504	424	424	424
Mean		273	162	153	270	285	430	928	433	39	295	630
Sigma		356	202	177	157	298	353	624	492	1120	276	387
Range		1-2906	1-994	1-975	1-546	1-1878	1-1291	1-2028	1-1999	1-818	1-1313	2-132
Runs*	1	79/99	135/156	78/113	115/116	116/115	58/67	93/116	78/86	76/98	86/92	89/96
Up/Down	2	44/50	70/69	41/41	39/46	45/39	28/29	52/38	58/48	46/32	43/43	46/33
	3	24/7	30/13	17/9	14/9	11/14	17/8	11/7	16/15	12/7	11/8	6/11
	4	9/2	7/5	14/2	4/2	1/4	5/7	6/3	3/4	4/1	4/0	3/3
	5	2/0	2/0	6/1	1/0	0/1	3/2	3/1	0/0	2/0	---	0/1
	6	1/0	0/1	4/1	---	---	1/1	---	0/1	0/1	---	---
	7	---	---	4/0	---	---	0/0	---	---	---	---	---
	8	---	---	0/0	---	---	2/1	---	---	---	---	---
	9	---	---	3/0	---	---	---	---	---	---	---	---
Log.	A	0.108	0.074	0.327	0.603	0.228	0.147	0.872	0.061	0.041	0.435	1.989
Sort	B	1.340	1.236	1.035	1.105	1.255	1.456	1.251	1.528	1.117	1.188	1.071
Fit	S	0.462	0.453	0.370	0.107	0.266	0.195	0.063	0.495	0.761	0.236	0.063
Auto-Regression		16(1)	3,5, 17(3)	3,5,11, 17(4)	2,3(2)	3,7, 23(3)	2,3,5, 17,19(5)	18, 23(2)	3,4,13, 19(4)	5(1)	3,11,14, 16,17(5)	3(1)
Kasiski		-(0)	7(1)	14,19(2)	-(0)	20,29, 33(3)	11,31(2)	13(1)	29(1)	-(0)	28(1)	7(1)
Modulo Reductions		2,4-6,8, 10,12,16, 20,22,24, 25(12)	2,4-8, 10,11, 13-25 (21)	5,10,15, 20,21, 25(6)	25(1)	-(0)	9,25(2)	25(1)	9,11,15, 18,20, 22,25(7)	4-25 (22)	25(1)	25(1)

Notes: * Run Up/Down of Length 7 or greater indicate significant deviations from a controlled or random process.

** Known solution with Beale's version of "Declaration of Independence" which has 1332 text words and 6527 letters.

which strengthens our conviction that B1 and B3 were written in a slightly different manner from that used to encode B2. Note that S1 and S2 bracket the three Beale Cyphers which gives us once more an indication that he used a process whose severity lies between the maxims used in these two computer simulated codes.

6.2 Summary of CRYPTT Data

Table 3 lists the results obtained from a Chi-Square analysis for letter frequencies of the various pseudotexts compared with standard English. The table has four major headings for the code data resulting from the three Beale Cyphers and the Poisson-generated random data. Each of the fourteen runs listed first in this table was set up identically to produce various pseudotexts from the stated keytext and the respective Beale Cypher, or the random code PS. Under Method I we created output using the word-counting method and went through a range of lags (or leads) from -5 to +5, for a total of eleven pseudotexts from the first letters of these words. We also produced outputs for second, third, etc. letters but they did not add anything new and are not shown. Under Method II, we used last letters and leads ranging from 0 to 2 for a total of three pseudotexts; again, second-last and other letters are not shown here although we did produce them. Under Method III we used the letter-counting method, ignoring spaces and punctuations, with lags (or leads) ranging from -10 to +10 for a total of 21 pseudotexts per run. Finally, under Method IV we used again the letter-counting method, including also spaces between words but ignoring punctuations, with a lag (or lead) ranging from -2 to +2 for a total of five pseudotexts. Thus all entries in Table 3 reflect averages over the number of pseudotexts produced by any of these four methods.

For reference purposes, the left side of the table also lists the languages of the keytext; the respective run numbers were also shown in Table 1. The first fourteen lines of the table show individual results for each text; later segments of the table provide information on certain averages. For example, line 1 of Table 3, referenced against Table 1 can be partially interpreted as follows: Run 1 against Beale Cypher 1 with the Latin Preamble to John's version of the Magna Carta, as listed in the Encyclopaedia Britannica produced eleven pseudotexts by Method I with an average Chi-Square deviation from standard Egnlish letter frequencies of 346. Method II with an average over three pseudotexts yielded a Chi-Square value of 698. The same document yielded 21 pseudotexts for Method III with a Chi-Square average of 175. A Chi-Square of 177 was finally obtained by Method IV for five pseudotexts.

Later segments of this table provide further averages on these Chi-Square values. AV 1-3 takes care of Latin versions of the Magna Carta with preamble. These averages are readily compared with Line 4 where we have a Latin text with the preamble omitted. AV 5-7 provides information about English versions of the Magna Carta including its preamble, while AV 8-12 summarizes results obtained for English versions under omission of the preamble. In this segment of Table 3, AV 13-14 summarizes results obtained from other English texts, namely Beale's version of the Declaration of Independence and a modern English text; both of these texts were chosen on

TABLE 3

CHI-SQUARE SIGNIFICANCE TESTS ON LETTER FREQUENCIES FOR FOUR DECODING METHODS

	Lan-	Beale Cypher No. 1				Beale Cypher No. 2				Beale Cypher No. 3				Poisson Random Data			
Line	guage	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
1	L*	346	698	175	177	555	1082	374	308	487	970	248	249	252	451	151	12
2	L*	380	775	173	142	590	1213	372	348	519	1015	237	240	321	646	145	15
3	L*	362	658	192	150	617	1154	368	386	571	946	248	245	333	498	153	16
4	L	721	587	165	171	1329	905	299	449	951	826	193	182	564	558	146	12
5	E*	370	479	128	103	676	868	302	190	545	724	193	146	285	348	93	8
6	E*	368	511	99	104	634	757	266	210	525	659	145	148	349	416	71	5
7	E*	461	532	126	118	810	997	297	209	614	872	192	180	405	496.	88	6
8	E	289	439	74	61	461	719	160	172	341	529	106	91	270	365	65	5
9	E	338	448	81	77	644	804	172	195	491	571	113	93	302	395	71	6
10	E	269	360	79	68	465	561	180	138	368	507	120	105	278	360	67	6
11	E	367	365	74	62	725	667	165	183	549	464	92	84	310	329	55	4
12	E	346	356	75	62	730	609	168	184	539	449	93	82	301	278	54	4
13	E	355	347	54	62	514**	624	147	147	462	418	71	68	278	315	50	5
14	E	305	236	61	45	532	464	166	149	416	388	92	84	286	248	43	4
1-3	AV	363	710	180	156	587	1150	371	347	526	977	244	245	302	532	150	14
5-7	AV	400	507	118	108	707	874	288	203	561	752	177	158	346	420	84	7
8-12	AV	322	394	77	66	605	672	171	174	458	504	105	91	292	345	62	5
13-14	AV	330	292	58	54	523*	544	156	148	439	403	82	76	282	282	46	5
1-4	AV	452	680	176	160	773	1088	353	372	632	939	232	229	368	538	149	14
5-12	AV	347	407	85	76	619	707	203	178	485	558	122	108	306	355	66	5
1-14	AV	377	485	111	100	663**	816	246	233	527	667	153	143	324	407	90	8

Notes: * With Preamble

** Known Solution in this Group with X^2=73 purpose to provide a basis for further benchmarks.

All Latin basic texts are averaged once more under AV 1-4 while all English texts are averaged out under AV 5-12. The last line of this table provides a grand average of all observed Chi-Square values. While there may be some question about normalization of the results of these Chi-Square tests between various keytexts, i.e., reading Table 3 from left to right, vertical comparisons can always be made without introducing any correction factors. Statistically speaking, this correction factor would involve division by the number of actual (e.g., not blank) data points in each cypher. A similar problem of normalization arises between methods.

Even a cursory examination of Table 3 reveals at once striking differences between Latin and English texts, Av 1-4 and AV 5-12, respectively . In all cases, Latin texts produce significantly poorer letter frequencies in the pseudotexts. It turns out that the derived letter frequencies are equally poor with or without the preamble. Therefore, we can conclude with a very high degree of confidence that the keytext was not Latin. In itself, this answers one of our earlier questions and contributes somewhat to the inferential knowledge we now possess about the Beale Cyphers.

Now we shall turn to a comparison of methods. Basically, Methods I and II refer to word-counts while IIIand IV refer to letter-counts. Let us first rule out Method II which uses last letters of numbered words. Not shown in this table are results obtained for second-last, third-last, etc. letters but they exhibit a like pattern. With few exceptions, Method II produces consistently results which are significantly worse than Method I, allowing us to discontinue any further consideration of Method II.

Now, there remain the original Beale word-counting Method I and two letter-counting Methods III and IV, without or with spaces between the words counted. As a benchmark, we have a Chi-Square level of 73 for the known solution to Beale Cypher 2, as indicated in a footnote to the table. We also have a benchmark from line 14 which introduces, on purpose, a keytext which could not possibly have been known to Mr. Beale. Focussing our attention on Av 5-12 for all English Magna Carta Texts, we find that Method I over Method III provides a 3.05 reduction in Chi-Square averages which is due to the change in number of observable data elements, for Beale Cypher 2. However, for B1 and B3 comparable reductions of 4.09 and 3.97 are obtained, while the PS random data yield a ratio of 4.64. Significant changes in respective ratios are observed for AV 1-14, but not for AV 1-4 and some others. Thus we have detected a structural difference between B1 and B3 on the one hand and B2 on the other. This difference is both language and data dependent. Using the random data PS as a benchmark, we find that B1 and B3 differ significantly from the expected reduction level. This conviction is further enhanced if we look at comparable values for AV 13-14 which contains two English texts but not the Magna Carta. Therefore, we now look only at line 13 and find that B2 reduces by 3.5 while B1 and B3 reduce by 6.58 and 6.50 respectively, while the random data PS yield only a coefficient of 5.55. This is the desired clue: For at least one document (Beale's version of the Declaration of Independence) Method III produces significantly better letter frequencies (albeit still garbled in the pseudotext) than expected, for Beale Cyphers B1 and B3, by comparison with Method II for B2.

A similar analysis of Method IV does not yield the same kind of significant change in comparisons between the random data PS and B2 versus B1 and B3. This observation rules out Method IV which we had only included for the sake of completeness. When setting up our programs we had never assumed that the cyphers' author would go to all the trouble of counting letters and spaces and/or punctuation marks. Such techniques have only recently assumed a dominant position because electronic data processing devices can handle such chores more efficiently than man.

7. Conclusions

The solutions to Beale Cyphers 1 and 3 have remained undiscovered despite time and the tenacity of many analysts. We would certainly have gone into deep shock if one of our CRYPTT runs had come out in cleartext and provided us with the geographical coordinates of the alleged treasure! In fact, had such been the case it is doubtful that this paper would have been written. Rather, the entire group responsible for this project would have shouldered pick and shovel and taken off for the Virginia hills. Nevertheless, the results obtained from this simulation study have contributed greatly to a better understanding of these cyphers and produced what we consider significant results.

By way of summarizing our findings, the following are statements of facts, as of this date:

(i) Beale Cyphers 1 and 3 are "for real." They are not random doodles but do contain intelligence and messages of some sort. Further attempts at decoding are indeed warranted.

(ii) The method used for encoding cyphers 1 and 3 is similar to that used for cypher 2. It is very probable that a letter-counting, rather than a word-counting method was used. If it was indeed a letter-count, then spaces between letters and punctuation marks were not included in the count.

(iii) The basic text used for encoding cyphers 1 and 3 is not Latin. There is enough evidence to assume it was English. Other languages have not been subjected to the type analysis indicated here.

Thus we return to the drawing board, as it were, hoping that sooner or later someone will find the right text(s) with which to decode Beale Cyphers 1 and 3. We doubt very much that it was the Magna Carta (English version) as suggested by our Pennsylvania visitor. Others before us have already ruled out the Declaration of Independence. However, with the help of our Univac 1108 CRYPTT programs it should be a simple matter to establish a systematic procedure and to test routinely any and all documents of relevance. If and when the mystery is finally resolved, we may have to eat much crow in view of the statements made above. Much time has elapsed since Beale buried his treasure and many people must have passed THE spot. It is quite likely that upon locating the vault it will be found empty. Whether located accidentally or by breaking the code, the first successful treasure hunter is not likely to reveal his find. It is only the second successful treasure hunter that will surface and cause a great deal of disappointment ... except in the circles of professional cryptanalysts!

8. Bibliography

1. The Cryptogram, Official Publication of the American Cryptogram Association, 9504 Forest Road, Bethesda, Maryland 20014 (Sesame, May-June, 1968.)

2. Sir Arthur Conan Doyle, The Complete Sherlock Holmes, Doubleday & Company Inc., New York, 1953 (The Valley of Fear, pp. 903 ff.)

3. Encyclopaedia Britannica, Declaration of Independence and Magna Carta.

4. Helen Fouché Gaines, Cryptanalysis, Dover Publications Inc., New York, 1956

5. Carl Hammer, Private Communications, Beale Cypher Study Committee, 2121 Wisconsin Avenue, N. W., Washington, D. C. 20007.

6. Frances Beale (Smith) Hodges, The Genealogy of the Beale Family; 1399-1956, Ann Arbor, Michigan, 1956

7. P. B. Innis, The Beale Fortune, Argosy, August 1964.

8. David Kahn, The Codebreakers, The MacMillan Company, New York, 1967

9. Al Masters, Has the Beale Treasure Code Been Solved? True Treasure, September-October 1968.

10. William Stubbs, Select Charters of English Constitutional History, Oxford at the Clarendon Press.

11. William F. Swindler, Magna Carta - Legend and Legacy, the Bobbs-Merrill Company Inc., New York.

INTERCOMMUNICATION

DIVISION OF SPERRY RAND CORPORATION

TO: Carl Hammer
Washington, D. C.

FROM (NAME): G. E. Mellen

LOCATION & DATE: St. Paul - 3 October 1967

DEPARTMENT: Nike Systems Division

SUBJECT: BEALE CIPHER

CARBONS:

Thank you for your prompt reply to my request for a copy of your Beale papers. I have not yet had the opportunity to look at the cipher in detail, but here are a few first impressions.

1. Since the right margin of the Argosy text is not justified and there are other typographical irregularities, is there some significance in the layout of the numbers? I will send to Roanoke for photostats of the originals; this is the only way to judge. (I will copy you on these.)

2. Reduction modulo 26, searching for cyclic additives, etc., seem too sophisticated in light of the alleged circumstances of time, place, and writer. Initially, I think I will assume the same general system of Message 2, the same conclusion you reached in your correspondence with Kahn.

3. Intuition is not always bad. Note the opening sequence:

71 194 38 1701 89 76 11 83

and assume the opening text:

T H E V A U L T

Then, on a printed page containing 200 words, the expected number of words having these letters as initials are, respectively:

36.26 9.5 5.54 0.82 24.72 2.82 2.48 36.26

Thus, all letters except "V" are reasonably likely to occur on the first page, and the jump to serial 1701 may perhaps result from a random searc of the next 10 pages.

22 SEP 1970

4. One may assume that the key text is at least 2911 words long. This may be misleading in that the author may have numbered only the first 3000 words of a much longer text. The name "Thomas Jefferson Beale" certainly suggests possibilities, but [illegible] after all these y
is there a chance some of TJ's papers remain undetected? [illegible] Probably not. Also, a man looking ten years ahead would not likely use the current Farmer's Almanac. Unfortunately, in Virginia, he might well use a classic Greek or Latin text under the assumption that Homer would always be Homer. The number of variorum editions of the classics rules out practical solution if this was indeed the case.

5. Based on the distribution of the numbers, my hunch is that the clear was written out, and then a few letters here and there were enciphered from the same page of a previously numbered key. If this supposition is true, it is not impossible that the typography of the key text can be reconstructed from assumptions of type size, leading, line width, fonts, and set widths. This in turn may lead us to a study of early 19th century Virginia printers--but this is far in the future.

6. What size truck do you think we will need?

Greg Mellen
G. E. Mellen

GEM/vrm

UNIVAC
CORPORATION
FEDERAL SYSTEMS DIVISION
2121 WISCONSIN AVE. N.W., WASHINGTON, D.C. 20007 • TEL. (202) 333-8

15 May 1970

MEMORANDUM #5

"The Beale Cypher" by Carl Hammer

Coincidence or not - the Beale Cypher Study Committee held its second and very productive meeting on the day of the solar eclipse, 7 March 1970. Thanks to the enthusiastic participation of all present, we were able to shed a considerable amount of light on the subject and were in no way detracted by celestial phenomena or gastronomical indulgences. This is by way of thanking all those who attended for their continued interest and contributions. Some day soon we will indeed get to the bottom of this mystery if not the bottom of the "excavation or vault six feet below the surface of the ground" where TJB allegedly deposited his fabulous treasure.

The Committee has grown beyond all expectations and there appears the need for continuing our past record of documentation and consolidation. Therefore, we are mailing out to all members on our list (and a few whose name will be added shortly) the following documents:

(1) Minutes of the second committee meeting, prepared impeccably and diligently by Dr. David L. Dobbins.

(2) Committee Membership List of May 1970.

(3) Reprint of "True Treasure," September/October 1968.

(4) Reprint of "Saga," March 1970.

(5) Copy of "CRYPT" Program Listing and Output.

(6) Copy of the last slide of my Beale Paper presentation: "I was only kidding."

Our archives have also been swelled through the addition of the following documents; copies should be requested from the respective authors (shown in parentheses and on the Committee Membership List):

(a) Tom Beale's Cave of Gold, CLIMAX Magazine, June 1958 (Bob Williams; a few copies also available from this writer until supply which Bob so kindly sent me is exhausted).

22 SEP 1970

2606 N. Brandywine Street,
Arlington, Virginia 22207

27 November 1968

Frances Beal Smith Hodges,
2170 Virginia Drive,
Wichita Falls, Texas

MRS LUKE HODGES
2164 VIRGINIA, Apt. "C"
WICHITA FALLS TEXAS 76309

Dear Mrs. Hodges:

In connection with my research on U. S. Western History for the period 1815 - 1825, I will be most appreciative for your assistance. In your "Genealogy of the Beale Family" you report Thomas J. Beale as having gone to Missouri. From other sources it appears that a Thomas Jefferson Beale and a party of thirty men made a number of hunting and business trips along what was to become the Santa Fe Trail.

On the first of these trips, the party is reported to have departed St. Louis on 19 May 1817 reaching Santa Fe about the first day of December where they put up for the Winter. Major trips with heavy wagons were reportedly made from Santa Fe to St. Louis into Virginia and return during the Summer of 1819 and the Fall of 1821. If verified, the Beale Party would be among the very first to use the old trail.

Does your background file contain references to Thomas Jefferson Beale and his hunting party activities? In particular, I am interested in copies of letters to, from, and about Mr. Robert Morriss of Lynchburg, Virginia who was believed to be a close friend of Beale and his party. Also, I am interested in the names of others in the Beale party.

If any of the above information is available in your background file or in any other file of which you are aware, you will be most helpful by advising me of it and of the cost for reproduction and mailing of any information that you may have. Enclosed is a self-addressed stamped envelop.

Sincerely Yours,

Carl W. Nelson, Jr.

Mr. Nelson- I do wish I might send you exactly the information you wish but unfortunately I have nothing on Thomas J. Beale beyond the fact that he was a son of Richard Eustace Beale.

Most of this family data was originally sent to me by Mrs. Julia Beale Renaker of Lexington, Ky., (d 1937) who was a niece of this man. I do not even know that J. stood for Jefferson.

It would be interesting to know the result of your research IF you have the time to so advise at that date.

Good luck to you

Frances Hodges

2000 N. Brandywine Street,
Arlington, Virginia 22207
11 December 1968

Dear Dr. Hammer:

This by way of bringing you to date on my activities relating to Thomas Jefferson Beale will answer some of Mr. Hammell's questions put in his fine letter of 31 Oct.

The time available for my work has been limited to study of materials that have com to me via the Virginia State Library. I have not had the time necessary to work on the map problem posed by Hammell. I may find time for a visit to the Library of Congress shortly after the first of the year. I have given a very careful reading to the Lynchburg, Va Press as follows:

May 1809 with breaks to 26 December 1816,
21 February 1817 to 24 April 1818,
21 September 1818 to 30 September 1819,
13 November 1819 to 3 October 1820, and,
20 April 1821 to 5 April 1822.

A check of the Hart Papers will show that the above periods coincide with:

Organization of the Beale Party ca April 1817,
The first gold shipment ca December 1819,
Beale's three month visit with Morriss ca January 1820,
The second gold shipment ca Fall 1821, and,
Beale's last visit with Morriss in January 1822.

In addition to a general search of these files, the five periods of special interest hav been checked by two detailed readings of the newspaper.

This study clearly confirms Robert Morriss' financial plight resulting from his speculation beginning ca November 1818(the Lynchburg Commodity Market listed tobacco bid 12 dols., asked 17 dols.). Apparently Robert Morriss executed deeds of trust for loans based on about 10,000 acres of Virginia tobacco land, on all of his Lynchburg property, and on his personal home, money from which allegedly was used in the tobacco futures market. His wife Sarah was a party to the deeds. By 8 April 1819, the Lynchbur market prices on new tobacco were 6 dols bid and 9½ dols asked, at least that is the way I read their market reports. Thus, Morriss had quite a parcel of high price contracts for tobacco to make good. The paper reported that many Virginians had lost out in such speculation. From here on the record clearly shows a series of foreclosure sales on all of Morriss' land holdings ending up with sale of a Lynchburg house, personal property, and slaves on 29 April 1820. It is interesting that had Morriss carefully evaluated reports on the European tobacco market which appeared in the Lynchburg Press during Sept./ October 1818, he might have avoided his losses. At any rate, this clearly confirms the story that we have from the Ward papers about Morriss' financial problems.

Robert Morriss advertised his business activities. At no time through 5 April 1822 during the periods cited in the Beale letters and in the Hart papers is there an advertisement by Morriss concerning a hotel business in Lynchburg. There were many hotel and tavern ads by other owners(A Mr. Hoyle owned the Franklin). At no time was there mentio of a "Washington Hotel". Thus, one may take Beale's words literally(9May 1822 letter), "Ever since leaving my comfortable quarters at your house . . . ". Beale may have been a guest in Morriss' home in 1822 as well as ca January 1820.

I have found no reference to a Thomas Beale in the Lynchburg Press during the perio of interest or otherwise. Possibly because this was a period when dueling was still in vogue, all letters to the editor type correspondence was signed in pseudonym. Because context, I am fairly certain that Robert Morriss used "Publicola" on his letters. While many letters from and about Missouri Territory were published in the Lynchburg Press, no can be tied to T. J. Beale. Because so many of the Beale Party came from around Richmond, I hope to check those papers for the same periods.

My next objective is a survey of National Archives files on the Department of State neral correspondence for the period April 1816 through December 1823. This material in microfilm at the National Archives. If Beale had an official assignment, some tip should appear in these records. Also, I plan a look at the records and orders for tary convoy assignments West from St. Louis for the period of interest. Further, I in process of gathering information on the Planters Hotel in St. Louis(according to ale's letters, he lived at this Hotel); and, eventually, I hope to have a look at the . Louis papers for the period of interest.

As is clearly evident, nothing has been added to our knowledge about the very elusive omas Jefferson Beale. The attached copy of my letter to Frances Beal Smith Hodges and r reply will show that she too has no additional information on TJB. Hopefully, the sults may be better as the study progresses. Whatever the outcome, I find this reading oject to be absolutely fascinating. I have given very little time to cypher analysis a result.

Cordially,

Carl Nelson
11 December 1968

: R. A. Hammell, ACA,
198 Richey Avenue,
West Collingswood, N. J.

BEALE CYPHER STUDY COMMITTEE

The second meeting of the Beale Cypher Study Committee(BCSC) was called to order by Dr. Carl Hammer, Chairman BCSC, at 9:10 AM 7 March 1970. The meeting was held in Dr. Hammer's office in the Univac Building on Wisconsin Avenue in Washington, D. C. Members and guest attending were:

Dr. Carl Hammer	(1), (2), (3)
Dr. David L. Dobbins	(1), (2), (4)
Mr. Fred Chesson	
Mr. Ken Fawcett	
Mr. Fred Steffens	
Mr. Carl Nelson	(1), (2)
Mr. Frank Speh	
Mr. Robert A. Hammell	(1), (2)
Mr. William Dansie	
Mr. Ralph Berger	
Mr. Robert E. Caldwell	(1), (2)
Mr. Robert N. Williams	
Mrs. Agnes Williams	

(1) Original Committee Members
(2) Present During Meeting Of 9/12/68
(3) Committee Chairman
(4) Committee Secretary During Meeting Of 3/12/70

After much persuasion , Dr. Dobbins agreed to act as cetary of the 7 March 1970 BCSC meeting.

Dr. Hammer passed around his collection of correspondence received during the past year. A few of the letters were amusing and one in particular "demonstrated" to the members the uselessness of continued work since a genius from Mass. had already solved the cypher.

DISCUSSION

Dr. Hammer began the discussion with portions of his last memo to members being considered, in particular, the computer work. He reported that to date he must have used 100 hours on the Univac 1108. He has used 3000 lags in the Declaration of Independence (DOI) out of a possible of some 6000 letters to determine a starting place without success. He plans to continue this work out to the end of the DOI.

Dr. Hammer stated that to his knowledge he was the only one to date in the BCSC to make money on the cypher--he was awarded the $500 prize at the recent ACA meeting in Florida.

In some of the programs run by Dr. Hammer since the Florida meeting he has shown that the Chi Square for all three of the BC's was reduced going from word count to letter count in the DOI. He reported, however, that the Chi Square was reduced

(b) Probable letter substitutions for cyphers 1 and 3; while in themselves inconclusive, these represent good efforts and should be looked at (Robert Caldwell).

(c) Miscellaneous Items of Interest; a list of eleven historical events and some bibliographic references (Robert Williams).

(d) Lynchburg and its neighbors (3 pages); The Genealogy of the Beale Family (3 pages); Thomas Beale & Celeste Grand Pierre (1 page); Colonial Families (2 pages); Colliers Magazine of 19 March 1927 (1 page); five synoptic reviews of books and articles, chuck-full of interesting and relevant information (Robert Hammell).

(e) Historical and Analytical Studies (21 pages); Beale Cypher Test Key (10 pages); two studies of substance which the author distributed (with a caveat) by mail to earlier committee members (Carl W. Nelson).

Communications received since our 7 March meeting include letters or notes from the following:

(i) Stan Czarnowski - Sorry about your car breaking down. Hope you have recovered fully and can resume digging for the Beale Treasure soon.

(ii) Koshka - Use your own judgment in the use of our TJB materials for ACA.

(iii) Carl E. Kesler - Your very encouraging note was read with interest by all who attended the last meeting; by all means let us know who "cracked" the cypher and we can then "bury" this Committee where the treasure was...

(iv) G. E. Mellen - Beale might have numbered his letters (or words) in multiples of five only, thus the noted preference for moduli 5, 10, 15, etc. This writer experienced similar problems in his own simulation efforts.

Last but not least, I want to say thanks to R. N. Williams for his loan to us of True Treasure and Saga. With his help we have enriched our archives and also those of the committee members. I am sure that all of us are grateful for the continued cooperation that so many have given to our efforts.

Happy solving to all and best wishes for a successful summer!

Sincerely,

Dr. Carl Hammer.

Carl Hammer, Ph.D.
Director
Computer Sciences

emu
att

much more for cyphers #1 and #3 than for #2, thus strongly implying to him that cyphers #1 and #3 are coded on letter count rather than work count as is cypher #2. Mr. Fawcett gave reasons at this point for his thinking that cyphers #1 and #3 were based on the same method of coding--word count.

HISTORICAL

After returning from a short break, Mr. Nelson began his discussion of the papers he had mailed to the members several weeks prior to the meeting. He stated that the proprietary notice placed on the papers was to be taken as "use the material at your own risk of being incorrect". He suggested that a person wishing to use the material should give him a call.

Mr. Nelson stated that he had proof that Dr. John Hamilton Robinson was working for Pres. Monroe--Monroe sent a letter in which he fired Robinson. Robinson is one of the possible candidates for TJB. He still thinks that TJB used a pseudonym. Mr. Nelson is still looking through the imcoming and outgoing Dept of State mail files. He stated that the military record sources are practically unproductive due to the first military orders for men in the area of Sante Fe were not until 1829. He is continuing to check papers from Shawnee Town, Missouri--the logical crossing point for TJB's party. Mr. Nelson stated that the Catholic Church kept good records of help given to U. S. citizens of the early 1800's. He plans to search these files that are presently held by LC.

Mr. Nelson suggested that we give consideration to Wards writings and also Thomas Jefferson's papers on Univ. of Virginia. He also suggested we search records of the following places for the passing of the Beale party:

a) Charleston, W. Virginia
b) Huntington, W. Virginia
c) Louisville, Kentucky
d) Morganfield, Kentucky
e) Shawnee Town, Illinois

He plans to check the Columbia Herald from Franklin, Missouri for the mention of Beale's party.

The next speaker on the program was Mr. Frank Speh. He suggested the group take a trip to the area where Beale supposedly buried his treasure. No action was taken on this suggestion. Mr. Speh thinks the first few words of cypher #1 are:

"To find cache go to ridge east of gap---"

and that the last word is:

"---rocks".

Next, Mr. Fawcett discussed positions of cypher #1 that he thinks are most vulnerable to trial substitution. The main position is possibly at 359 and 464.

Next, Mr. Hammell discussed some of the Beale history in terms of genealogy studies he had performed. A copy of his works is included with this package.

Next we launched into a facetious discussion of how to get the gold out once we break the code. Any suggestions? No action was taken!

CONTINUATION OF WORK

We next discussed work for the coming year. The following items were discussed:

a) Continue historical studies--Mr. Nelson would be focal point for this.

b) Dr. Hammer plans to write a program to substitute words and phrases into cyphers and have output presented on scope face in his office. This program will be of the real-time interactive type--person sits in front of scope face, assigns a word and appropriate letters are displayed at candidate locations; words can also be shifted. This program should be finished by mid-summer.

c) Continue using key text to try to decode cyphers. Refer to Dr. Hammers last communication to find what has been tried already.

d) No one in the group present had ever seen anything on the Beale cypher except for that material sent out by the Richmond Library and certain magazine articles. Someone should try to see the original material if possible, or the earliest transcription.

MATERIAL TO BE REPRODUCED

Certain material was left with Dr. Hammer to be copied and distributed. If anyone has material he thinks important to the group, he should send it to Dr. Hammer to be copied.

The meeting was adjourned at about 1:15 pm for eclipse watching.

After watching the eclipse, several members of the group had lunch together up the street from the Univac Building. At lunch Dr. Hammer made a suggestion which sounds worth further consideration. His suggestion was to form a Beale Cypher Study Committee Corporation. Such a corporation would cost the members only a small sum (probably less than $10.00), and would be beneficial in many respects such as,travel and expenses related to corporation related work. The main benefit to most of us would be in the areas of tax deductible expenses. I would encourage the members to give much consideration to this point.

These minutes are respectfully submitted by:

David L. Dobbins

David L. Dobbins

HOW DID TJB ENCODE B2?

Carl Hammer

[Originally presented at Beale Cypher Symposium, Washington, DC, 15 April 1972]
In the context of this study, it is quite immaterial whether the Beale Cyphers are a hoax or not. The question of authorship and authenticity of the three cyphers is properly the subject of another investigation. Here we are concerned only with the question how B2 was encoded since both cleartext (B2C = "I HAVE DEPOSITED...") and keytext (DOI = Declaration of Indepenence, Beale Version) are known to us.

Having decided on this particular method of encoding the (B2C) cleartext by using first letters of numbered words in the keytext (DOI), how would a schooled cryptographer proceed? The answer depends on such factors as the desired coding efficiency, time available, and support materials at hand. Given much time and ample quantities of paper and pencils, an expert cryppie may want to begin by making a complete list of all available keytext letters so that he can develop a cypher with a maximum number of degrees of freedom. For example, B2C requires 42 A's but DOI can furnish 167 words starting with that letter. Therefore, our cryptographer could encode every letter A in B2C with a different number. In fact, he can make his selection in many permutative ways. On the other hand, the DOI has only two V's but to encode B2C we need eighteen of them; at best, we must use each V in DOI nine times. Finally, as is well known, there are no X's and Y's available in the DOI and our cryptographer must make some provisions to cope with this deficiency.

TBJ, or whoever authored the three cyphers, has given us a clue what to do about the last question by encoding X = 994 ("Sexes") and Y = 822 ("Fundamentally"). We observe casually that a better pun for the latter choice would have been the word "by" which occurs sufficiently often at the thirteen position numbers 90, 221, 577, 682, 706, 755, 852, 957, 1020, 1030, 1069, 1106, and 1198. Naturally, we would not want to use these elements for B but that would not cause any problems; DOI provides us with 48 B-elements but B2C requires only twelve of them. Thus, we would still have an abundance of B's, even if we were to use up nine of them for the punned Y's.

Evidently, the author of B2 was not at all that clever. Additionally, he chose most of his substitutions from the lower numbered elements and repeated them rather unnecessarily. For example, instead of using a differently numbered element for each occurrence of the letter A, he chose a subset of 15 from the whole set of 167 available elements, repeating some of them as often as five times. An analysis of the code elements used indicates that a majority of them come from the beginning of the DOI: 407 below 100; 574 below 200; 637 below 300; 673 below 400; 690 below 500; 718 below 600; 732 below 700; none used between 700 and 800; and 818, 882, 994 used for V, Y, and X, respectively.

A modern cryptographer - or even an experienced one of the early nineteenth century - desiring to encode B2C with the DOI as keytext in a most efficient manner would, therefore, have gone through the following steps:

(1) Develop a masterplan for this project.

(2) Prepare a listing of required letter frequencies for the B2C cleartext.

(3) Prepare a listing of available letter frequencies in the DOI keytext.

(4) Analyze the letter frequency ratios (see table I) for available-to-required letters and decide which course of action to follow in four cases:

TABLE I

Letter Frequency Distributions

λ	DOI		B2C		E	$R\left\{\frac{[n(\lambda)_{DOI}]}{[n(\lambda)_{B2C}]}\right\}$
	n(λ)	f(λ)	n(λ)	f(λ)	f(λ)	
A	167	.1267	43	.0564	.0781	3.9
B	48	363	11	144	128	4.4
C	53	401	17	223	293	3.1
D	37	280	49	642	411	0.8
E	36	273	105	1376	1305	0.3
F	64	484	21	275	288	3.0
G	19	144	15	197	139	1.3
H	80	606	37	485	585	2.2
I	68	515	55	721	677	1.2
J	10	76	2	26	23	5.0
K	4	30	1	13	42	4.0
L	34	257	32	419	360	1.1
M	29	220	6	79	262	4.8
N	19	144	69	904	728	0.3
O	144	1075	63	826	821	2.3
P	63	477	12	157	215	5.2
Q	1	8	-	-	14	-
R	40	303	38	498	664	1.1
S	64	484	48	629	646	1.3
T	254	1923	70	919	902	3.6
U	28	212	25	328	277	1.1
V	2	15	18	236	100	0.1
W	59	447	13	170	149	4.5
X	-	-	4	52	30	0.0
Y	-	-	9	118	151	0.0
Z	-	-	-	-	9	-
	1323	1.0001*	763	.9999*	1.0000	

* Due to roundoffs not exactly 1.0000

Notes: DOI Refers to the first letters of the 1323 words in the Beale version of the Declaration of Independence.

B2C Refers to the cleartext obtained by decoding Beale Cypher 2 with the Declaration of Independence as keytext.

E Refers to standard English letter frequency.

Obvious
suspect
hypothe
did not
iar wit
ments f
in spur
(i) the
"distan
using "
use "di
limited
of his

What e
179 di
below 1
pointe
are be
the au
always
IHAVED
elemen
657 ev
others

Table
atitio
differ
I = 14
result
road t
to rep
tition
in add
a new
the la
air a
coded
dicat

(4.1) More elements available than required for certain type letters: Develop a strategy of random selection without duplication;

(4.2) The same number of elements available as required for certain type letters: Develop a more tightly controlled strategy for random selection, still without duplication;

(4.3) Fewer elements available than required for certain type letters in the cleartext: Develop a strategy for randomly repeating the use of the available letters, with a uniform distribution for repeated usages;

(4.4) No elements available for some letters required (examples: X, Y): Develop a clever substitution strategy so as to minimize repeated usages (examples: X = Sexes, Y = By).

Obviously the author of B2 did very little in this direction. Rather we may suspect that he "learned by doing," and we propose, herewith, the following hypothesis: B2 was encoded in a grossly suboptimal manner, because the author did not care to produce a most efficient code and was probably not even familiar with such a concept. Rather, he set up a limited list of available elements for the first several hundred words of the DOI and added to this list, in spurts, as he proceeded with his encoding task. He also noted quickly (i) the paucity of V's in the DOI and he decided to use consistently the "distant" element 818, (ii) the absence of X's and he decided to "pun" it by using "distant" element 994, and (iii) the absence of Y's and he decided to use "distant" element 822. All these "distant" elements are far beyond the limited range of available elements he had "planned" to use for the balance of his encoding job.

What evidence do we have in support of this hypothesis? First, B2 contains 179 different numerical elements (after removing repetitions), 80 of which are below 100, another 36 between 100 and 200, and only 63 are above 200. As we pointed out earlier, most of the repeated elements (75 percent, to be exact) are below 200. Details are shown in Table II which gives the order in which the author picked "new" elements (obviously, the first element in every row is always "new") to encode his cleartext. For example, to encode IHAVEDEPOSITED..." he chose I = 115 (rather than I = 2) for his very first element; but he gets around to using I = 2, 8, 67, 115, 140, 154, 158, 314, 657 eventually. Yet he fails to employ I = 139, 151, 165, 167, and many others which occur "early" in his numbered keytext.

Table III analyses the pattern he employs to introduce new numbers against repetition of elements already used. The first four "I's" are encoded with all different, new numbers 115, 657, 140, 2; but 657 is a surprise! The next I = 140 is a repetition; it is followed by three new numbers 8, 154, 314. The resultant pattern (details not shown here) indicates that halfway down the road the author decided that he had enough different numbers and no longer had to replenish his stockpile. Typically, the last 33 I's encoded are all repetitions of numbered elements used earlier. A surprise element in this pattern, in addition to an occasional use of a very high number, is the introduction of a new element toward the very end of his encoding process: the last three B's, the last T, the last U needed to encode B2C were seemingly picked out of thin air and for no apparent reason. On the other hand, the last 76 E's are encoded by re-using earlier elements! A weighted density plot would clearly indicate his preference for a seemingly lazy and very inefficient process.

TABLE II

	Order of Elements by First Occurrence in B 2		Counts		
	In Order of Appearance	Sorted	E	O	T
A	24 36 28 177 45 81 98 51 284 150 27 229 83 25 152	24 25 27 28 36 45 51 81 83 98 147 150 152 229 284	7	8	15
B	308 9 77 18 134 495 193	9 18 77 134 193 308 495	3	4	7
C	84 65 92 4 94 199 21	4 21 65 84 92 94 199	3	4	7
D	52 15 210 118 63 252 135 246 320 406 591	15 52 63 118 135 210 246 252 320 406 591	7	4	11
E	37 49 7 79 85 138 190 629 496 520 557 612 584 33	7 33 37 49 79 85 138 190 496 520 557 584 612 629	6	7	14
F	195 159 122 273 131 360 676 11	11 122 131 159 195 273 360 676	3	5	8
G	270 48 113 133	48 113 133 270	2	2	4
H	73 107 394 6 20 301 204 466	6 20 73 107 204 301 394 466	5	3	8
I	115 657 140 2 8 154 314 158 67	2 8 67 115 140 154 158 314 657	6	3	9
J	120 590	120 590	2	-	2
K	305	305	-	1	1
L	42 101 102 233 400 157 196 420 176 405	42 101 102 157 176 196 233 400 405 420	6	4	10
M	58 82 117 207	58 82 117 207	2	2	4
N	47 10 287 353 616 549 44 566	10 44 47 287 353 549 566 616	4	4	8
O	31 56 5 136 46 106 12 43 57 125 143 302	5 12 31 43 46 56 57 106 125 136 143 302	6	6	12
P	17 105 30 121	17 30 105 121	1	3	4
Q			-	-	
R	59 53 96 219 248 344 112	53 59 96 112 219 248 344	4	3	7
S	62 35 71 78 110 38 216 515 609 297 242 283 275	35 38 62 71 78 110 216 242 275 283 297 515 609	6	7	13
T	22 29 26 543 3 41 16 34 60 61 14 50 32 64 39 653 288	3 14 16 22 26 29 32 34 39 41 50 60 61 64 288 543 653	10	7	17
U	238 316 95 250 371 388 409 440	95 238 250 316 371 388 409 440	5	3	8
V	818	818	1	-	1
W	72 290 19 66 40 1 459	1 19 40 66 72 290 459	4	3	7
X	994	994	1	-	1
Y	882	882	1	-	1
Z					
			95	84	179

NOTE: E = Even Counts, O = Odd Counts, T = Total Counts

TABLE III

Alternate Appearance of New and Used Elements in B2

	N	U	N	U	N	U	N	U	N	U	N	U	N	U	N	U	N	U	SUMS
A	4	1	2	2	2	1	4	1	1	4	1	3	1	16					43
B	3	2	1	2	3														11
C	3	2	3	1	1	7													17
D	2	1	3	7	1	4	1	3	1	1	1	4	1	6	1	12			49
E	5	2	2	3	1	5	2	4	2	1	2	76							105
F	3	4	1	2	1	1	1	1	1	1	1	4							21
G	3	2	1	9															15
H	2	1	4	5	1	18	1	5											37
I	4	1	3	3	1	9	1	33											55
J	2																		2
K	1																		1
L	8	8	1	11	1	3													32
M	3	1	1	1															6
N	4	5	1	1	2	6	1	49											69
O	6	2	3	6	1	7	1	7	1	29									63
P	3	1	1	7															12
Q	-																		-
R	2	1	1	2	1	1	1	1	1	4	1	22							38
S	4	1	1	4	2	1	1	4	1	1	1	13	1	1	1	7	1	3	48
T	9	2	1	1	1	3	1	10	1	1	2	13	1	23	1				70
U	4	1	2	13	1	3	1												25
V	1	17																	18
W	6	5	1	1															13
X	1	3																	4
Y	1	8																	9
Z	-																		-
																			763

NOTE: N = New, U = Used

Summarizing our findings, we observe that our author proceeded substantially in the following way:

(1) First, he wrote out his cleartext B2C; we may surmise that he edited it at least once before he began with the encoding process.

(2) He then numbered (at least) the first 1000 words in the DOI, noting (in passing?) what elements to use conveniently for X, V, and Y.

We state here specifically that the author did not make a complete list of all available letter elements which would have been more efficient and also speedier. We also assert that he affixed position numbers to each of the circa 1000 first words of the DOI.

As shown in Table II, of 179 different numbers used, 95 are even and 84 are odd. A statistical test reveals that the ratio 95/84 does not differ significantly from 90/89 or 89/90! Thus, there is no preference of evens over odds, nor any other bias.

(3) He then encoded his cleartext, jumping randomly from one section of the keytext to another. These jumps were noted in our earlier paper (1) when we observed significant periodicities of lengths 3 and 5, as well as significant deviations between runs up and down. A casual inspection of B2 is quite convincing in this matter; if we single out elements belonging to specific hundreds, B2 assumes the following appearance: 115 - (73, 24) - 818 - (37, 52, 49, 17, 31, 62) - 657 - (22, 7, 15) - 140 - (47, 29) - 107 - (79, 84, 56) - 238 - (10, 26) - etc. This simplistic notation enhances our ability to see the pattern which was basic to the author's encoding job.

Finally, we observe that TJB was certainly not beyond making many clerical errors in his encoding process. We have several communications giving details about the forty-four specific code elements (numbers) which were edited by George Hart against the earlier Hiram Herbert papers in an effort to "force" the solution. Specifically, George Hart applied corrections of plus-one to two elements, 53 and 84; of minus-one to twenty-one elements ranging from 96 to 241; of plus-ten to four elements ranging from 449 to 505; of plus-nine to eleven elements ranging from 511 to 620; of plus-ten to three elements 643, 647, 666; of plus-eleven to two elements 807, 811; and of minus-eleven to the largest element 1005.

Anyone who has ever tried to number words by hand in a given keytext will at once recognize these corrections as being of a very common type. Incidentally, in the ranges shown above are also some few numbers which required no corrections suggesting that "Uncle TJB" went about his task rather sloppily, to say the least.

Unfortunately, we have no Wellsian Time Machine to help us "see" the author of these cyphers at work. But we have the tools of analysis and simulation both of which point strongly in the direction of our original assertion: TJB was not a professional cryptographer! However, he must have had some exposure to the coding techniques employed in his time; in his hour of need he resorted to this type of multiple substitution (homophonic) cypher. His choice has the advantage of great relative security which lasts exactly as long as the keytext is unknown. After that, even a high school student can decypher any such codes. TJB botched his job rather badly, making numerous mistakes in the numbering of the words and in the selection of clever substitutes for missing letters. He also did not guard himself too well against attempts to break the

cypher by
of it. C
text ele
B1 and B

The latte
uals and
will they
its of g
tions" we
massive f
tools of
can write
lic domai
of our g
the surfa

1. Aaron,

2. Gaines

3. Hammer
Commun

4. Hart,
the Ro

5. Herber
tions,

6. Hodges
1956)

7. Hohman
March/

8. Innis,

9. Kahn,

10. Knigh
Janua

11. Maste
Sept.

12. Matya
Multi
of Io

13. McGeh
93-12

14. Shafe
Priva

15. Sinko

16. Thaye
Treas

17. Uffel
Ciphe

cypher by probable word or letter substitutions, probably never even thought of it. Otherwise he would have used a greater diversity of available key-text elements to obtain a maximum in degrees of freedom. May we infer that B1 and B3 also contain many such errors?

The latter two have withstood the rather disorganized attack by many individuals and groups successfully for well over a hundred years. How much longer will they be able to protect "the exact locality of the vault" with its deposits of gold, silver, and jewelry? Judging by the many rumblings about "solutions" we are inclined to believe that B1 and B3 will shortly yield to the massive force exerted by organized and determined experts employing the latest tools of modern technology. It will be interesting to see what epitaph we can write about them as the decoded messages B1C and B3C filter into the public domain and the press. It will be equally interesting to look at the faces of our golddiggers as they enter the *empty* "excavation or vault six feet below the surface of ground"...

REFERENCES

1. Aaron, Frank H., *The Beale Paper's History*, unpublished manuscripts.
2. Gaines, Helen Fouché, *Cryptanalysis* (New York: Dover, 1956).
3. Hammer, Carl, Signature Simulation and Certain Cryptographic Codes, *Communications of the ACM*, 14 (January 1971) 3-14.
4. Hart, George L., Sr., *The Beale Papers*, 67 page manuscript available from the Roanoke Public Library, Roanoke, VA, 1964
5. Herbert, Hiram J., P. O. Box 126, Roanoke, VA 24002, Private Communications, 1970
6. Hodges, Francis Beale (Smith), *The Genealogy of the Beale Family (1399-1956)*, Ann Arbor 1956.
7. Hohmann, Robert E., Beale Code No. 3 Deciphered, *True Treasure*, March/April 1973, 23-25.
8. Innis, Pauline B., The Beale Fortune, *Argosy*, August 1964.
9. Kahn, David, *The Codebreakers* (New York: Macmillan, 1967).
10. Knight, Gary H., The Beale Ciphers, private communication (unpublished), January 1978.
11. Masters, Al, Has the Beale Treasure Code been Solved?, *True Treasure*, Sept./Oct. 1968.
12. Matyas, Stephen M., *A Computer Oriented Cryptanalytic Solution for Multiple Substitution Enciphering Systems*, Ph. D. Thesis, The University of Iowa, 1974.
13. McGehee, Joseph W., The Beale Papers, *The Cryptogram* (December 1954) 93-120.
14. Shafer, Robert E., West Virginia Wesleyan College, Buckhannon, WV 26201, Private Communication, 31 January 1972.
15. Sinkov, Abraham, *Elementary Cryptanalysis* (New York: Random House, 1968).
16. Thayer, Victor I., Captain Beale's Lost $5 - $10 Million, *National Treasure Hunters League*, 4(2) 1974, 42-47.
17. Uffelman, Malcolm Rucj, *An Application of Unicity Analysis to the Beale Ciphers*, private communication (unpublished), February 1978.