BRINGING HOME THE CRYPTOGRAPHIC BACON

A. ROSS ECKLER
Morristown, New Jersey

For more than one hundred years various iconoclasts have speculated that Francis Bacon wrote Shakespeare's plays and sonnets. Some have searched for indications in the Shakesperean corpus that Bacon cryptographically hid some statement of his authorship. Most of the older claims were ably discredited by the eminent cryptanalysts William and Elizebeth Friedman in their 1957 book *The Shakespearean Ciphers Examined*. This article briefly examines a more recent claim of Baconian cryptography, Penn Leary's *The Second Cryptographic Shakes-Peare* (Westchester House Publishers, Omaha NE 1990). Leary, a lawyer and apparently self-taught Shakespeare scholar, used the digital computer to aid in his search for Baconian fingerprints, enabling him to sift through literally mountains of evidence. But did he sift too thoroughly? Did he, in fact, discover "Bacon" in random concatenations of letters?

Overcryptanalysis--the extraction of apparently-meaningful patterns from random data--is the trap that Baconians must strive to avoid. But it is not easy to specify when overcryptanalysis has occurred. The sticking-point lies in the following question: what is the universe of possible messages? what combinations of letters (in addition to the one discovered) would also have been credited as legitimate? If there are too many of these, the plausibility of the discovered message is correspondingly diminished. In certain instances, plausibility can be mathematically calibrated, as will be shown below.

Leary began his investigation with a relatively focused search for a Baconian signature--the 31 words and 146 letters of the dedication to the Sonnets:

> TO THE ONLIE BEGETTER OF THESE INSVING SONNETS Mr. W.H. ALL HAPPINESSE AND THAT ETERNETIE PROMISED BY OUR EVER-LIVING POET WISHETH THE WELL-WISHING ADVENTVRER IN SETTING FORTH
>
> T.T.

(the precise line-arrangement is not reproduced, nor are the periods between each word included).

Although Bacon wrote a short essay on the biliteral cipher, Leary elected to look for a Caesar cipher instead, one which shifts the plaintext alphabet a fixed number of letter (say, E to A, F to B, G to C, etc.). Trying "everything I could think of" (not very helpful for the precise specification of plausible alternatives!), Leary discovered that when the alphabet was shifted four spaces (as illustrated above), the

final letters of the five words "oF thesE insvinG sonnetS mR" were transformed into BACON!

Could this have happened by chance? Using data compiled byh Fletcher Pratt in *Secret and Urgent: The Story of Codes and Ciphers* (Blue Ribbon Books, 1942) for frequencies of terminal letters in running text, one finds that Pr(F ending) = .049, Pr(E ending) = .223, Pr(G ending) = .225, Pr(S ending) = .137 and Pr(R ending) = .048 which, multiplied together, produce a satisfactorily-low random probability of FEGSR equal to .0000018.

But wait! It's time to consider other messages which Leary would have credited Bacon with inserting in the text. To begin with there are 27 such sets of letter endings (OEERF, EERFE, etc.) and 23 other alphabet shifts (the Elizabethans combined I with J and U with V), each with its own probability of generating BACON (CBDPO, DCEQP, EDFRQ, etc.). Secondly, there is the matter of alphabets. Leary indicates in his book (page 144) that he considered at least 12 (those with ending letters TVW, TVW, TVZ, TZV, TVX, TXV, TVY, TYV, TVXY, TVYZ, TVWY, TVZX). These made no difference in the appearance of BACON that Leary found, but they would have made a difference if other alphabetic shifts were considered. Thirdly, what about the previously-alluded-to "everything I could think of" (page 15)? He mentions testing first letters of individual lines, last letters, first and last, fourth letters of individual words (FORTH in the dedication might be a clue), second letters of individual words (note TO), and last letters of individual words (which resulted in success). In other parts of his book he tests consecutive letters (TOTHEONLIE...), alternate letters (TTENI... or OHOLE...), etc. And, finally, Leary elswhere considers reversed spellings (i.e., encrypting the name NOCAB). All these alternatives increase the probability of discov-ering BACON somewhere in the message.

How much? The probability of any of the 24 different Caesar encryptions is .0000029; curiously, Pr(FEGSR) accounts for well over half the total, which might have inspired Bacon (if he were aware of terminal-letter probabilities and the multiplicative rule) to use this as the most-easily-accommodated Caesar cipher. (Bacon's task--by no means trivial--was to insert a name-cipher by modifying the dedication without using awkward spelling or phraseology which would arouse reader suspicion that something strange was going on.) Since there were 27 sites where the cipher could appear (words 1-5, 2-6, etc.), the probability further increases to .000078, and reversal doubles this to 0.00016. Assuming 20 different plausible alphabets, the probability finally becomes .0032.

To factor in the "everything I could think of" one must must carry out analogous calculations using different letter-probabilities. The simple letter probabilities in running text compiled by Fletcher Pratt were used (pr(A) = .082, pr(B) = .014, ...), although chained bigram frequencies would introduce greater verisimilitude. In any event, the overall probability of the FEGSR shift and its 23 allies leads to a similar value,

.0000035. Multiplying this by 142 sites (TOTHE, OTHEO, ...), reverse encryption and 20 alphabets, one obtains .0000035(142)(2)(20) = .020. The same number is obtained if one looks for alternate-letter encryptions, every-third-letter encryptions, etc. First-letter-of-each-word encryptions have not been calculated, but probably are like last-letter ones. So, adding the probabilities for last-letter, first-letter, every letter and every-other letter--would Leary have also considered yet other letter-patterns as grist for his mill?--one obtains a probability of 2(.0032) + 2(.02) = .046 that BACON would have shown up even if not deliberately inserted by Bacon. This probability, interestingly, is about equal to the standard confidence level employed by professional statisticians to decide on the reality of an effect (such as the efficacy of a medicine vs. a placebo). So--the conclusion at best is "maybe".

Having demonstrated the BACON cipher to his satisfaction, Leary discovered other messages before it--a phonetic spelling (BEKAAN), an allusion to ciphering (CYPPHRS, a spelling not sanctioned by the OED), and an allusion to John Napier, the inventor of logarithms (NYPIR). It is difficult to formulate rigorous alternatives to these, but I find it difficult to believe that Bacon was so enamored with Napier that he would have added this information to his putative self-identification cipher.

The next discussion will illuminate Leary's propensity to overcryptanalyze. However, one should recognize that this overcryptanalyzation does not invalidate the analysis above; it seems likely that Bacon, a busy man, would have had time to carefully design the insertion of only one, or at most very few, cryptographic signatures. It would be unfortunate for posterity's search for evidence if Bacon were so inept in selecting his cryptographic signature that it could be frequently mimicked by randomly-occurring combinations of letters.

Leary's overcryptanalysis results from his conviction (first aroused by the previously-mentioned BEKAAN) that Bacon was wont to encode his name phonetically in the plays of Shakespeare. It is true that Elizabethan orthography was somewhat shaky, but Leary is extremely generous in his allowance for alternative spellings. Dozens of different examples are cited in the later chapters of his book, enabling the statistician to formulate a reasonably-rigorous set of "acceptable" alternatives as far as Leary is concerned. The first letter must be B, the second and fourth letters can be one or two choices from the set AEHIOVY, the third letter can be one or two letters from the set CKQ, and the fifth letter must be N. A flavor of the possibilities, including some outside these rules: BAQCAEN, BIHCAIN, BIQIN, BYIKEAAN, BIYHICAN, BAICVQEN. Would Bacon have really sanctioned all these spellings?

The probability of all these alternatives--there must be at least 37,000 of them--can be readily calculated from multiplying individual-letter or letter-group probabilities of the enciphered letters. Using a four-step Caesar substitution cipher as before, and the ABCDEFGHIKLMNOPQRSTVY alphabet, BACON is TSVKI. The probabilities are Pr(T) = .093, Pr(one or

two of SAEDKQR) = .524, Pr(one or two of VFM) = .093, and Pr(I) = .073, leading to a product of .000173, which is doubled for reversals to .00035. There are almost exactly one million words in Shakespeare's plays; if the average word-length is five letters, this means that phonetic BACON can be found 1750 times using just one Caesar cipher and one Elizabethan alphabet. This number must be increased if one uses other letter-patterns (beginnings of words, every other letter. etc.) but the message is clear--BACON is findable on every page! One must doubt that Francis Bacon went to <u>that</u> much trouble to cite his authorship!

THE PALINDROMIST

Mark Saltveit, PO Box 471258, San Francisco CA 94147 is the editor and publisher of a delightful new zine on palindromy, available for $12 per year ($20 overseas). He plans to publish four issues per year; the first two, dated Fall 1996 and Winter 1996, have already appeared. The first issue contains an in-depth biography of the Greek author Sotades, reportedly the originator of palindromic writing (although no examples by him survive), as well as the discovery of a little-known 1950s-era palindromic board game called Noggin; the second features political palindromes and the work of "amazing palindromist" John Connett, author of a couple of 1996 Word Ways articles. Saltveit encourages reader contributions and in fact promises a free copy of the issue in which one's palindrome appears. A reader challenge: write a clear and understandable palindrome containing a word with the trigram GHT (an example: NIGHT: FIFTH GIN). There's even a cartoon by Jon Agee!

Computer aficionados can reach Mark Saltveit through e-mail: palindromist@realchange.org. The address of his Web site is http://www.realchange.org/pal; this contains high-quality palindromes plus material that couldn't fit in the magazine.