

Eric Sams and Julian Moore

Cryptanalysis and historical research

Times Literary Supplement, 4 march 1977

In 1641 John Wilkins, a young chaplain later to be Bishop of Chester and a founder of the Royal Society, published the first English textbook on cryptography, *Mercury, or the Secret and Swift Messenger, showing how a Man may with Privacy and Speed communicate his Thoughts to a Friend at any Distance*. Most of his proposals were more ingenious than usable; but his book was timely. Within months of its publication the Civil War broke out and enciphered writing was used on a more extensive scale than ever

before. Wherever the keys to these ciphers have been lost or mislaid, modern historical research may be frustrated.

Some fitful progress has proved possible: for example, where a contemporary decipherment yields a key which is also found to fit other texts in the same cipher which have not survived *en clair* (as with some of the letters of Queen Henrietta Maria published by M. Green in 1857). Similarly, access to one text *en clair* may permit the decipherment of a whole correspondence, a possibility to which contemporaries themselves were alive. Thus when the Parliamentary forces captured the drafts of some Royalist cipher letters, they were soon able to publish the most discreditable passages from other such correspondence. There is now, however, little likelihood that the archives will yield similar windfalls for the modern researcher.

But the spread of cryptography stimulated its countermeasure – cryptanalysis, that is, the methodical application of techniques which permit decipherment without the key. In modern times this has become first mechanized and then computerized in ways well beyond the technical or financial resources of the average historian. But far simpler methods can yield effective and even impressive results. Thus in the nineteenth century a few dedicated full-time cryptanalysts made important contributions to the elucidation of state papers. But historical cryptanalysts are today a rare if not extinct species; and there is evidence that historians both in Europe and in America have now dug down in the archives to an apparently impenetrable bedrock of documents in cipher.

So the answer to the problem of unsolved ciphers still left in the archives is for research workers to acquire the relevant skills and to do it themselves. Uninstructed personal endeavour in this field must often have led to total frustration, or at best to disproportionate delay. A recent example is the cipher diary kept by Beatrix Potter between 1881 and 1897, left unstudied until 1952 and not solved until 1958 (by Leslie Linder) - although a competent cryptanalyst could have broken the system in about half an hour.

This article offers general advice on the compilation of the basic equipment which the historian will need should he set out to elucidate an enciphered document. It takes a whole library to encompass the history and practice of cryptography. But it happens that one particular cipher-system dominated European diplomatic and military correspondence for several centuries. The reason for its dominance is itself instructive. In *The Advancement of Learning* Francis Bacon defined the main virtues of ciphers "whereby they are to be preferred" thus: "that they be not laborious to write and read" and "that they be impossible to decipher." But these aims were already antithetical; and it was precisely the fruitful tensions between convenience on the one hand and security on the other that gave rise to the use of number-cipher throughout Europe from the sixteenth century onwards.

Ciphers operate by obscuring the most easily recognizable characteristics of the written word. Decryption (i.e., decipherment without the key) relies on using those patterns of language which the encipherer has not destroyed to provide the foundation for a reconstruction of the plain text. Convenience of encipherment favours the method of substitution, in which each letter of the message loses its usual identity and is regularly replaced by another symbol. Classically the method may be simple and formular, e.g., for A write D, for B write E, and so on.

Suetonius said that such alphabetical displacements were used by Julius Caesar, and whatever the

truth of this story the results are still known as "Caesars." During the Middle Ages substitution more often involved the use of specially invented symbols; and the more cabbalistic these looked the better. But such systems were soon seen to be insecure because all letters in all languages behave in predictable and hence identifiable ways. Thus the most frequent symbol used would be that representing E, since this letter is commonest not only in English (and especially in *olde Englishe*) but also in French, German, Italian, and Spanish.

To render this substitution less vulnerable the encipherer has to go further than merely disguising the identity of the plain text letters. So extra obscurity may be added by: the allocation of more than one equivalent to each letter; what Bacon calls "intermixtures of nulls and non-significants;" additional symbols for two or three-letter groups or common words or proper names; the avoidance of division into separate words. While such devices are well designed to camouflage linguistic patterns, and particularly the letter frequencies, they require more symbols than the alphabet provides.

The practical alternative first adopted (for example in the diplomatic ciphers associated with the name of Sir Francis Walsingham) was to supplement the alphabet with a plethora of invented signs. Their often arcane appearance endeared them to nineteenth-century romancers, and variations were used by Poe, who employed a selection of printer's signs in *The Gold Bug*, and Conan Doyle, who adopted arbitrary patterns in *The Dancing Men*. But in real life the labour of inventing and writing unfamiliar symbols soon rendered Walsingham's picturesque systems obsolete. The only set of equivalents that can provide sufficient variety and yet is familiar enough to be conveniently written is the series of integers; hence the dominance of number-ciphers.

These systems may be made as simple and formular as "Caesars" for the sake of convenience (e.g., E may be 30, 31, and 32), or else randomized for the sake of security (e.g., E may be 3, 29, 45, 61, and 89). In either case, by the late sixteenth century it is usual (but by no means invariable) for a one or two-figure number to represent single letters while three-figure numbers stand for syllables, common words, or proper names. This three-figure usage, properly called code rather than cipher, is often arranged in alphabetical order and is to that extent vulnerable to analysis.

The cipher component (i.e., replacement of letters by substitutes) is in principle entirely vulnerable, given a sufficient length of message. As a rule of thumb, the convenient minimum length may be calculated as about twelve times the number of different symbols used; but texts where the ratio is much smaller will be accessible to the skilled analyst. Very few, if any, of these ciphers are, technically speaking, indecipherable - although insufficient material may render them so for practical purposes - and Poe's dictum applies to them as it does not to some modern systems: "It may well be doubted, whether human ingenuity can construct an enigma of the kind which human ingenuity may not, by proper application, resolve."

A wish to exercise such ingenuity is the prime qualification of the amateur cryptanalyst. All researchers have the powerful incentive of contributing to knowledge by being the first solver ever, with added bonuses of possibly significant discovery and personal satisfaction. Second (but only second) comes the right cast of mind. The latter was defined more than thirty years ago by the psychologists who advised on recruitment to the British cryptographic service, which played so vital a part in the Second World War. The selection system was designed to identify general linguistic and problem-solving ability enhanced by special aptitudes for, *inter alia*, mathematics, chess, crosswords, and orchestral score-reading.

All such skills can be acquired in some degree. Cryptography is essentially a discipline, despite its important intuitive aspects. Its basic principles have been lucidly set out in elementary and advanced cipher manuals. There has been little systematic treatment of the number-cipher here discusses, though several sources (a list is appended) offer relevant commentary)

* * *

The first essential is to know the cipher's background and provenance. Here the research worker has a clear advantage over the cryptanalyst because of his knowledge of the sources, including the prime clue of the probable subject-matter of the text in question. But he will need suppleness of mind to avoid the frustration of the *idée fixe*. As he becomes more familiar with material from one source, the researcher will be able to recognize and employ to his own advantage the foibles of the encipherer - his preference for certain systems, say, or his penchant for protecting the beginning and end of his messages with many nulls, perhaps leaving the main body of the text comparatively vulnerable.

The next step is to identify the language used. This knowledge can offer unexpected insights. Thus a

number-cipher in the hand of Abraham Cowley was used for two letters to Charles I, one from Baron Jermyn and the other from Henrietta Maria. (The cipher was not one of those elucidated in the standard edition of her correspondence.) The frequency and pattern of the numerals used in the two texts differed in such a way as to suggest that the same cipher-system, probably designed for use in English, was being used to convey messages in two different languages - English and French. One significant clue was an apparent affinity between several cipher numerals that could only have arisen from the need to encipher the letters Q and U. A French cipher system could be expected to disguise this common (in French) diagraph; but it is rare in English and the system available to Cowley accordingly did not provide any special camouflage. The tentative identification of these letters began the chain of reasoning that led to the recovery en clair of the French text, the much less accessible English text, and to the discovery of a key which has been found to fit other unsolved cipher correspondence of the period.

To identify such patterns and combinations it is always essential to count and analyse the whole cipher-text, as one of the preliminary steps, in as much detail as possible. Index cards provide a serviceable method.

Take as an example of analysis a phrase from a letter of the exiled Charles II intercepted by agents of the Commonwealth and published in the papers of Thurloe, Cromwell's chef de cabinet (State Papers, Volume 111, page 76): "Upon the whole matter let me heare from you 114.20.28.41.66.25.63.30.32.68.31.44.167, in such a manner as may at least fully instruct me of what I may looke for." Divide each index card notionally into a grid suitable for entering the numbers used (10 x 10 for the two-figure cipher, with space for three-figure groups as necessary, is a suitable framework). On the card headed 1 14 write 20 in the appropriate cell; on the card headed 20 write 28; and so on (using dots for repetitions) throughout the whole message. With increasing experience the practitioner can readily devise more complex and informative systems for intractable texts. The emerging patterns of frequency and juxtaposition, though no doubt obscured by the devices already described, are nevertheless likely to permit certain inferences - for example that two or more different numbers throw up patterns sufficiently analogous to suggest that they represent one and the same letter.

The card index system will suggest letter relationships which may then themselves be the subject of further close analysis, and the most unpromising looking feature can in this way become the fulcrum for a solution of resistant material. At a crisis in his career, during a quarrel with Charles I, Prince Rupert wrote an often quoted letter, the greater part of which is in cipher, to his trusted friend Will Legge, the governor of the Royalist stronghold Oxford. Rupert seems to have been at pains to keep most of his meaning secret and used his cipher system skilfully to break up the texture of his message to an exceptional degree. The systematic analysis of such letter relationships as repetitions and reversals has, however, recently led to the decryptment of the text.

The cipher manuals cited below will give many other examples of detailed analysis. As with Rupert's cipher, their general rules will need to be modified or supplemented (in ways far too detailed and varied to be dealt with here, even in summary) by experience and practice with the cryptography of a given country and period. In these circumstances the researcher is well placed to supplement analytical methods by inference from extant plain texts, for example, to try to find the "probable word" hidden beneath the Cipher.

Thus the Cipher letter from Henrietta Maria to Charles I referred to earlier was found to include her customary and touching salutation "mon cher Coeur", which usefully confirmed some tentative identities between cipher and plain text. In all contexts the process of inference, by hypothesis and test by cross-checking is crucial.

In the example from Thurloe's papers cited above, the reasoning (much simplified for the sake of brevity and cogency) might run thus: "Let the language be English. The look of the frequency count strongly suggests a formulal cipher, with consecutive groups of numbers representing letters in alphabetical order. Most frequent are the early 30s; try them as E. Hence E E 68 E. So 68 is probably a consonant. T seems plausible in itself and about the right distance down the alphabet. Then 63 and 66 might well be R and S; which would give S 25 R F E T E. So 25 looks like C; try D I S C R E E T E. That yields D for 28, which assorts Well with 25 as C. Then 20 looks like A; and if 41 is I, then 44 suggests K (bearing in mind that in the English seventeenth century I and J are, like U and V, identical). Then the request must be for "a discreet key", making 167 = EY; so the three-figure groups are probably also in alphabetical order and 114 will be BY (rather than IN or WITH)."

In practice this process was reinforced and counter-checked at every stage by the substitution of proposed equivalents into the main body of cipher text, with encouraging results. So the request was in fact an indiscreet betrayal of the key actually used. That degree of naivety in cipher and encipherer alike, together with the failure of the interceptors to decipher the text, may suggest a certain lack of awareness of security and its techniques both in Charles's Court and Thurloe's cabinet; and the study of historical cryptography might permissibly include analogous inference from the type and use of cipher-systems.

Similar analysis, and "probable word" formulae such as dates and subscriptions can also provide vital points of entry into the unread shorthand systems which lie in the archives alongside cipher. For example, the shorthand used by William Clarke as secretary to Cromwell and his generals in Scotland has been declared by stenographic experts to be unreadable. But a cryptanalytic approach proved more fruitful; a frequency count and other analyses enabled some 300 equivalents to be identified, so that the whole system is now in principle readable wherever it is legible. It turns out to be similar to shorthand used by Pepys; Shelton's *Tachygraphia* (six editions were published between 1620 and 1641) was no doubt the common source book.

The disciplines of cryptanalysis can be taxing, and the tasks arduous, but the expenditure of time and effort is hardly greater than that required by the more cryptic of crosswords, and there is always the possible bonus of genuine and fruitful discovery. Such results are surely well within the competence of the average historical researcher, who already possesses the background knowledge and has only to acquire some relevant technique and dispel some irrelevant mystique. The only other needs are photocopied holographs (never transcriptions), plenty of cards and paper, sharpened pencils and toils, technical information or experience, reasonable confidence, plenty of patience, and a modicum of good fortune.

* * *

Selected sources:

J. Schooling: "Secrets in Cipher" in *The Pall Mall Magazine* (1896) ; contains historical background and useful facsimiles)

E. Bazeries: *Les Chiffres Secrets Dévoilés* (1901); has examples of Napoleonic number-cipher.

A Meister: *Die Geheimschrift im Dienste der Papstlichen Kurie* (1906) has examples of papal number-cipher.

H. Gaines: *Cryptanalysis* (1939), still the standard work on methods of cryptanalysis.

F. Pratt: *Secret and Urgent* (1939) offers a technique for decrypting multiple substitution (page 65)

L. Smith: *Cyprography* (1943), an excellent primer.

D. Underdown: *The Royalist Conspiracy* (1960) contains an appendix of Royalist letters during the Commonwealth, with examples of deciphering.

D. Kahn: *The Codebreakers* (1968); the unabridged edition is by far the fullest and best history; its detailed bibliography is invaluable.

C. Carter: *The Western European Powers 1500-1700* (1971).

S. Richards: *Secret Writing in the Public Records, Henry VII-George II* (1974).

The solutions to the cipher letters of Henrietta Maria, etc (SP 106, 10 f unnumbered) and Prince Rupert. (SP 16.511. f.89 have been communicated by Eric Sams and Julian Moore respectively, with notes on method, to Dr Ian Raw, Kings College London (1976). Solutions to unsolved cipher in Thurloe by Eric Sams have been filed in the Bodleian Library with Clarendon MS 94 (1973) and the solution to Clarke's shorthand by Eric Sams with the Clarke archives in Worcester College Library, Oxford (1974).