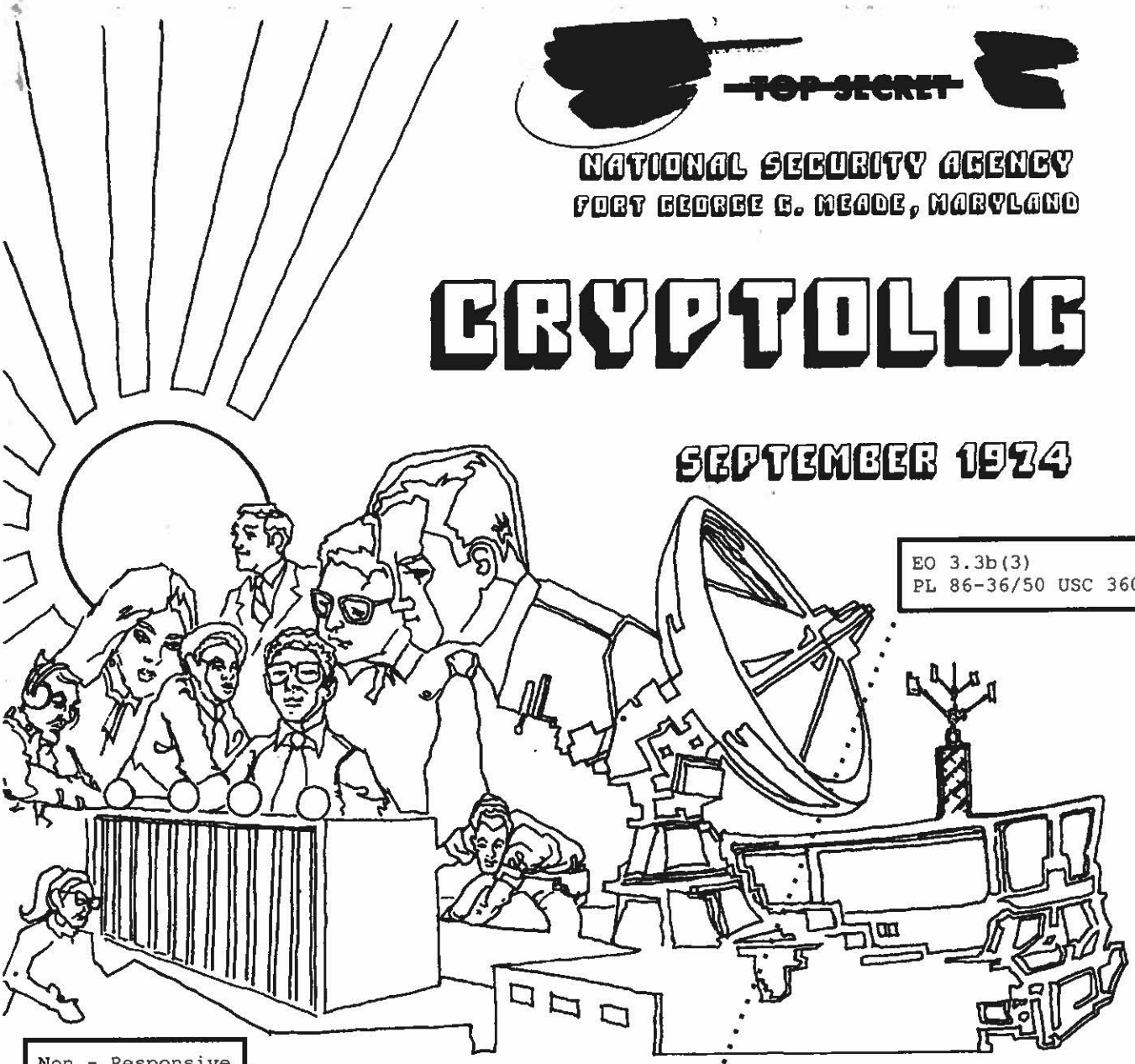
[redacted] TOP SECRET [redacted]

NATIONAL SECURITY AGENCY
FORT GEORGE G. MEADE, MARYLAND

CRYPTOLOG

SEPTEMBER 1974



EO 3.3b(3)
PL 86-36/50 USC 3605

Non - Responsive

[redacted]
Classified by DIRNSA (NSAM 123-2)
Exempt from GDS, EO 11652, Cat. 3
Declas Date Cannot Be Determined

THIS DOCUMENT CONTAINS CODEWORD MATERIAL

TOP SECRET

EO 3.3b(3)
PL 86-36/50 USC 3605

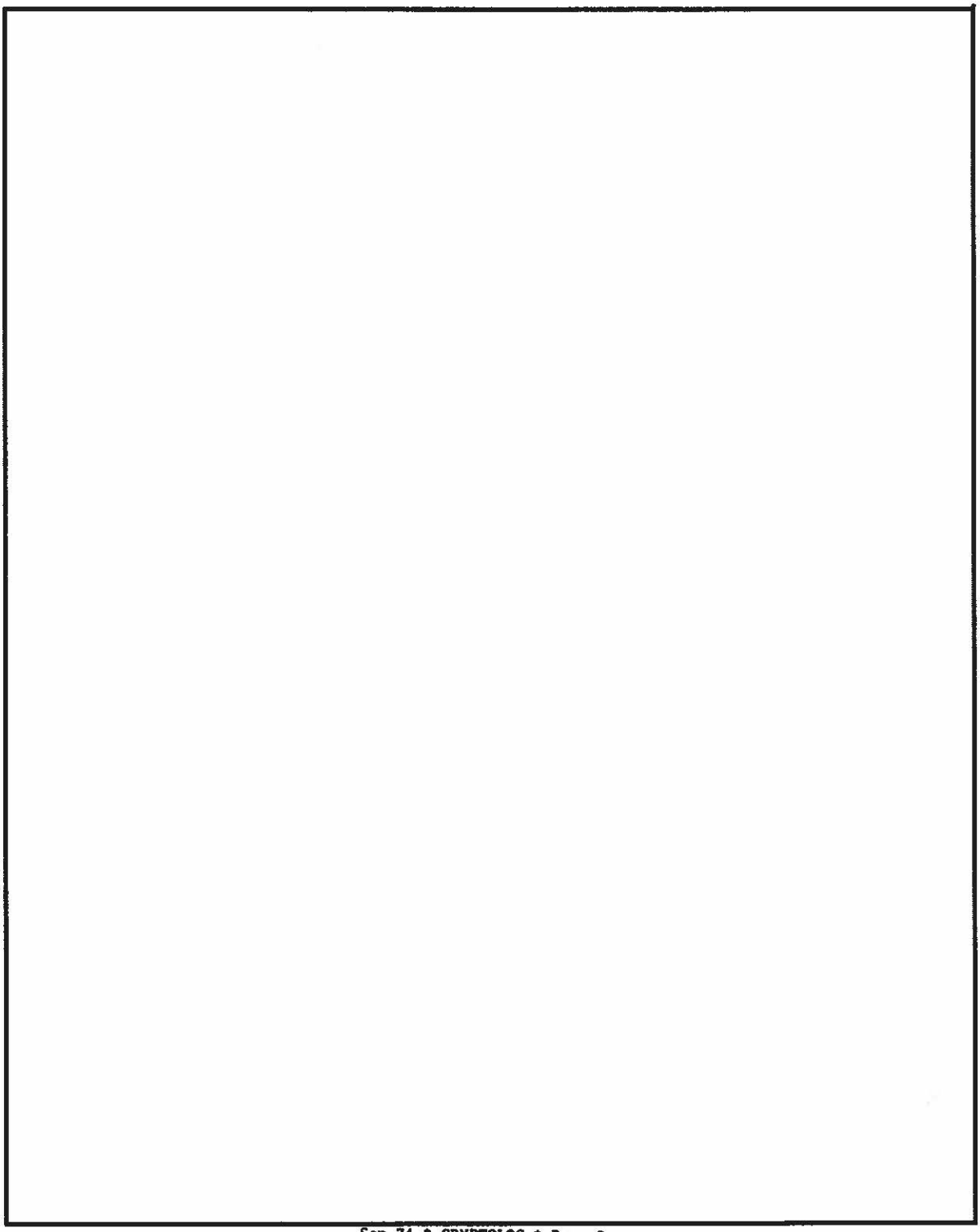
~~TOP SECRET UMBRA~~

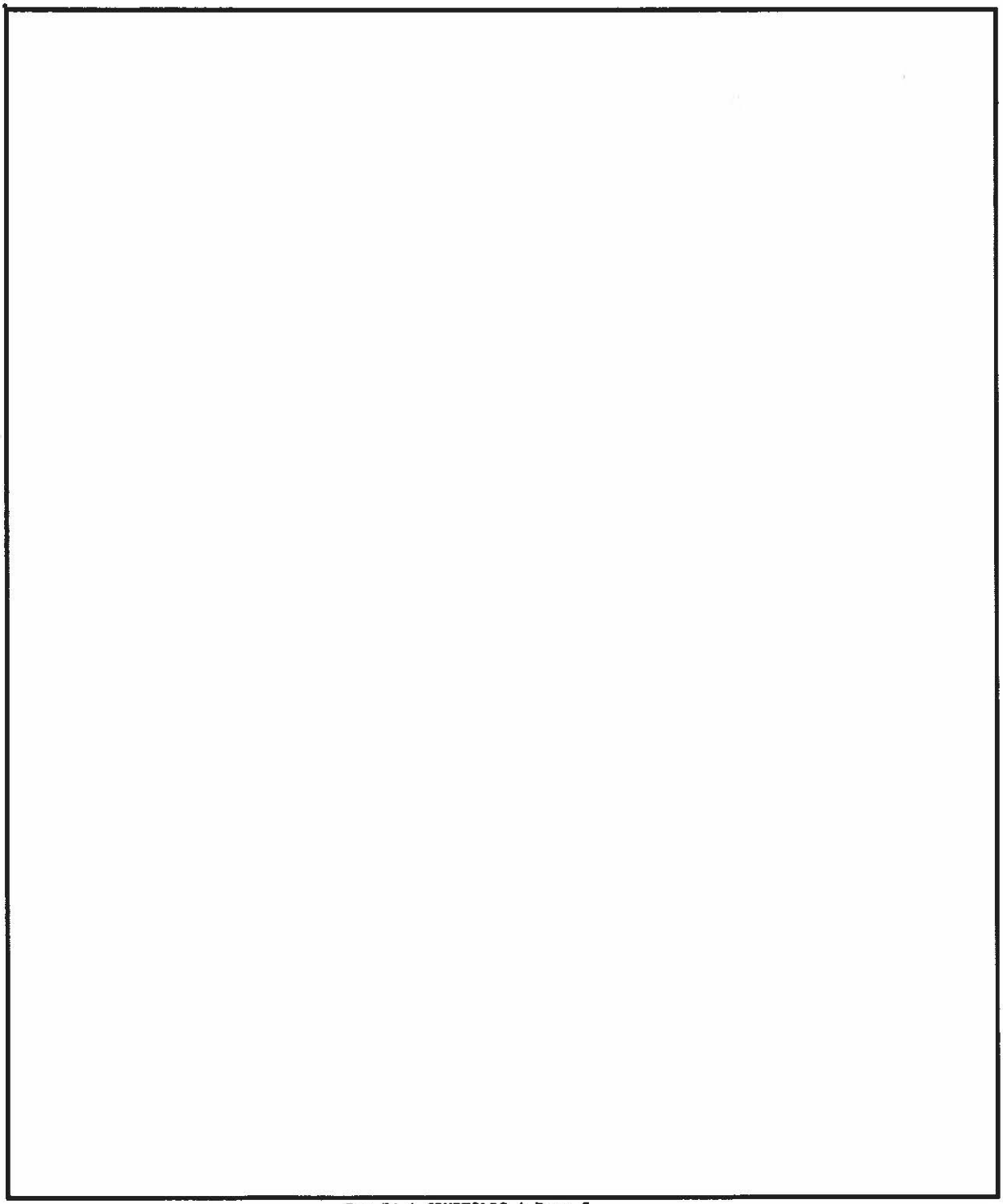
by DEREK K. CRAIG

EO 3.3b(3)
PL 86-36/50 USC 3605

~~TOP SECRET UMBRA~~

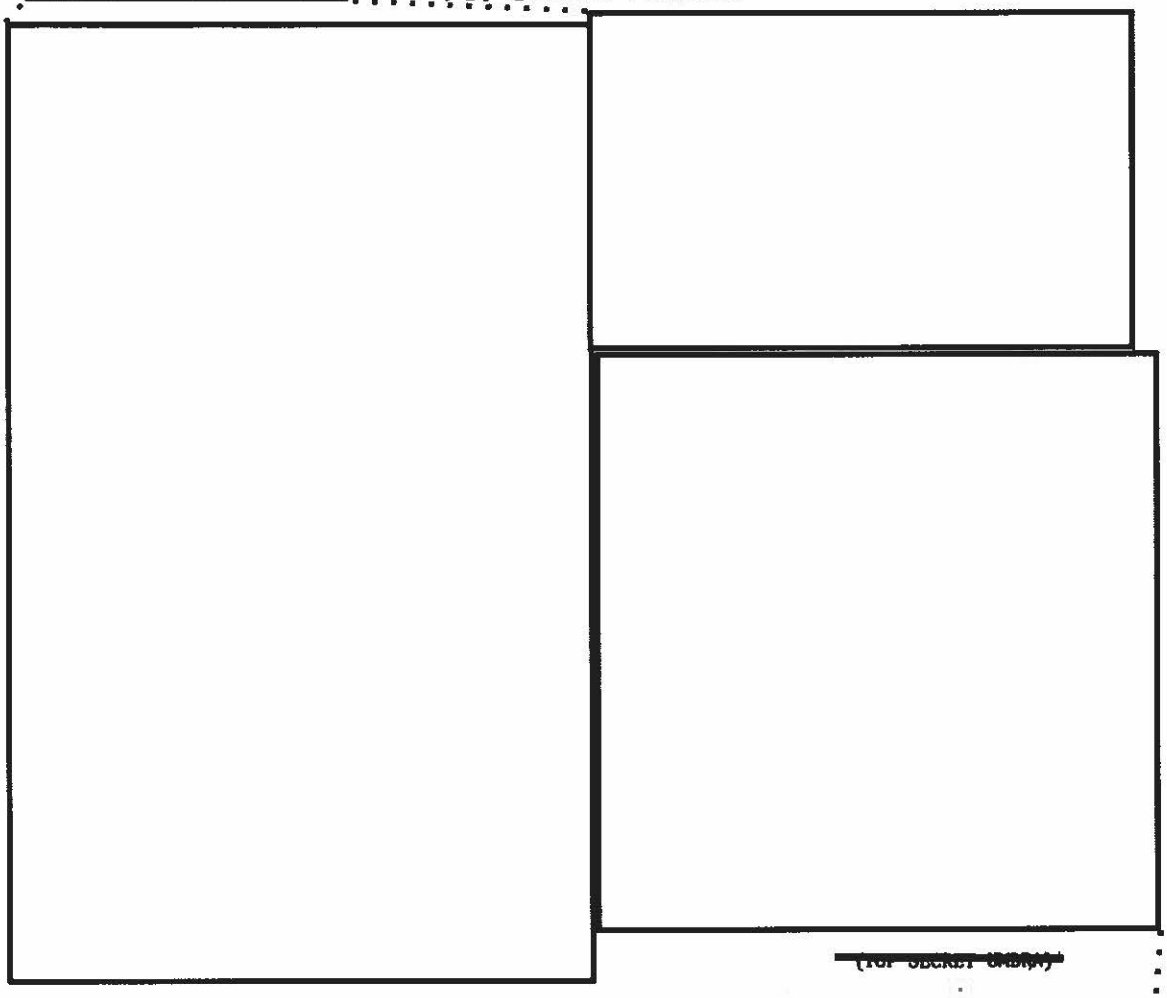
EO 3.3b(3)
EO 3.3b(6)
PL 86-36/50 USC 3605



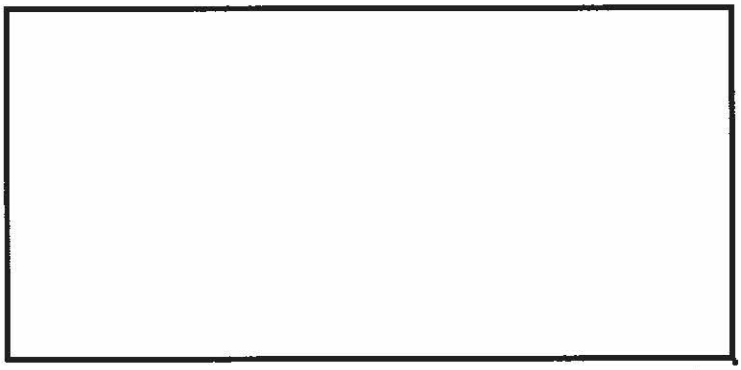


EO 3.3b(3)
EO 3.3b(6)
PL 86-36/50 USC 3605

~~TOP SECRET UMBRA~~



~~(TOP SECRET UMBRA)~~



~~TOP SECRET UMBRA~~

EO 3.3b(3)
PL 86-36/50 USC 3605

~~SECRET~~

CRYPTANALYSIS & CODE RECOVERY

by MARJORIE MOUNTJOY

The role of the cryptanalyst in code reconstruction overlaps that of the cryptolinguist, and both are often covered by the ambiguous term "bookbreaker". If we accept the refinement proposed by Stuart Buck, recognized Agency authority on code reconstruction, the cryptanalyst "breaks into" a code, while the cryptolinguist goes on from there to "reconstruct" the codebook.

We have some notes prepared several years ago by Marjorie Mountjoy, P1 cryptanalyst now retired, about what the cryptanalyst in such a case should know, have, or do to ensure that the whole process of code reconstruction can build on a reliable body of pertinent material. Here, with minor alterations, are the Mountjoy notes. They demonstrate once again that age is not a true measure of validity or value.

The cryptanalyst and the code reconstructor should work closely together. They may even be the same person, but, in any case, they should aim at compatibility, wherever possible. The compatibility is especially desirable when it comes to logging messages and requesting machine runs.

A tidy organization will desire the following tools (I say "desire," for they probably will not be permitted to store so large a collection):

1. Records of previous cryptographic practices.

[redacted]

Mr. Buck adds:

"Code reconstructors should always have easy access to these materials, especially where successor codes are concerned. The store of materials and information is important, not for the sake of hoarding, but to reduce the time required to exploit a new system. In this business, nothing is so wasteful as constant rediscovery of the wheel."

* * *

The following definitions and formulae are of utmost importance:

True base: The values assigned are known to be the same as those used by the original cryptographers.

Relative base: The values assigned have the same delta relationship as those of the original groups; i.e., they are off by a constant of 1111, 2222, etc.

Arbitrary base: The values assigned can be converted to the true by the addition of a constant amount which is purely arbitrary. (The Basic Cryptologic Glossary does not make this distinction between relative and arbitrary base. It should.)

~~SECRET~~

~~HANDLE VIA COMINT CHANNELS ONLY~~

Repeat rate of a code: $\frac{\sum_{i=1}^{n} f_i(f_i-1)}{N(N-1)}$

Delta I.C.: $\delta = c \cdot \frac{\sum_{i=1}^{n} f_i(f_i-1)}{N(N-1)}$

* * *

The following relationships will be extremely useful to the analyst working

* * *

Before turning the material over to the cryptolinguist for reconstruction of the codebook, the cryptanalyst is responsible for making sure, if possible:

1. that the code is not already recovered;
2. that the material is, in fact, plain code;
3. that the encipherment has been completely and correctly removed;
4. that the material is homogeneous;
5. that the code is reduced to true base (not always possible).

The five points may seem obvious, but I can cite some heartbreaking examples, as well as some heartening ones:

The material was reclaimed and reexamined. The monome-dinome hypothesis was quickly established, but not before the cryptolinguist had some agonizing moments.

* * *

Some procedures are simply not safe. One such procedure On the other hand, the cryptanalyst must have the courage of his convictions and be willing to defend his solution of a transposition bust which yields no repetitions at all.

I don't want to get into the details of additive solution, but the wise cryptanalyst will use statistical information to guide him to the shortest road to recovery. The repeat rate of the code (discernible from the depths) should tell him whether the solution can be obtained in a few days by hand methods or whether it may take a year or more and a full-scale mechanized attack to complete his part of the code solution. He should be able to spot a one-part code from the repeat rate of the beginning digraphs.

A resident code reconstructor is often desirable, even in the very early stages. may enable the cryptanalyst and the cryptolinguist to reinforce each other in unusual ways.

[redacted]

Code reconstructor and cryptanalyst may not see eye to eye on machine work, particularly on inverse frequency lists. The former often requests that the low-frequency groups be dropped, and the total gets lost. That is an unfortunate loss for the cryptanalyst whose best estimate of the probability of occurrence of a high-frequency group is $\frac{F}{N}$. The cryptanalyst needs such information to prepare log weights and to assemble lists of groups with a specified probability. He also needs it to compute the repeat rate. (Note: This problem is solved by the current practice of showing the relative frequency as well as the absolute frequency of a code group. Ed.)

Both the code reconstructor and the cryptanalyst rely on statistical methods, though the former sees them as only a small part of his store of tools. He is looking for a one-to-one correspondence between code groups and words or phrases, and he is guided by the frequencies of these items. (The measure of the extent of that guidance varies according to the individual's role and point of view--cryptanalyst or cryptolinguist.)[a]

The cryptanalyst is in more desperate need of statistical guidance. He is often in the position of a man looking for a needle in a haystack when the needle looks exactly like a piece of hay.

* * *

Virginia Valaki of G54 comments: "It is possible to have a 'tidy organization' with the large collection Marjorie described [redacted] in addition to general cryptologic files, a separate Code Library. The library is designed as a research center for codes: it is the repository for obsolete and surplus codebooks, runs and other code materials, the collecting point for statistics and reports on [redacted] a gathering place for general information on code reconstruction, machine support for bookbreaking and the state of the art; it is a wellspring of original studies and innovative procedures.... And it is possible to have comprehensive archives without bursting at the seams, thanks to microfilm and microfiche."

~~(SECRET/UMBRA)~~

[redacted]

[a]Mr. Buck's comments illustrate the difference in assessment of the role of statistics in code reconstruction. He says: "Code reconstruction is primarily a linguistic process, but it involves a lot more than mere language knowledge. The code reconstructor is, in fact, an analyst in the broadest sense of the word. Besides language fluency, he must have an intimate knowledge of the country under study, its history, culture, governmental structure, key personalities, geography--and a host of other matters. Codebooks are reconstructed through knowledge, not through ignorance. The aim of code reconstruction is to establish what is provable--not what is plausible."