[redacted]

~~Confidential~~

WAR DEPARTMENT

OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

MILITARY CRYPTANALYSIS

PART I

[redacted]

~~Restricted~~

WAR DEPARTMENT

OFFICE OF THE CHIEF SIGNAL OFFICER

WASHINGTON

MILITARY CRYPTANALYSIS

Part I

MONOALPHABETIC SUBSTITUTION SYSTEMS

By

WILLIAM F. FRIEDMAN

Principal Cryptanalyst
Chief of Signal Intelligence Section
War Plans and Training Division

PREPARED UNDER THE DIRECTION OF THE
CHIEF SIGNAL OFFICER

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON: 1938

30 April 1959

Paul S. Willard
Colonel, AGC
Adjutant General

MILITARY CRYPTANALYSIS, PART I. MONOALPHABETIC SUBSTITUTION SYSTEMS

Section		Paragraphs	Pages
I.	Introductory remarks	1– 3	1– 6
II.	Fundamental principles	4– 8	7–10
III.	Frequency distributions	9–11	11–17
IV.	Fundamental uses of the uniliteral frequency distribution	12–16	18–26
V.	Uniliteral substitution with standard cipher alphabets	17–22	27–39
VI.	Uniliteral substitution with mixed cipher alphabets	23–34	40–58
VII.	Multiliteral substitution with single-equivalent cipher alphabets	35–36	59–62
VIII.	Multiliteral substitution with multiple-equivalent cipher alphabets	37–40	63–69
IX.	Polygraphic substitution systems	41–46	70–98
X.	Concluding remarks	47–50	99–104
XI.	Appendix		105–120

SECTION I

INTRODUCTORY REMARKS

	Paragraph
Scope of this text	1
Mental equipment necessary for cryptanalytic work	2
Validity of results of cryptanalysis	3

1. Scope of this text.—*a.* It is assumed that the student has studied the two preceding texts written by the same author and forming part of this series, viz, *Elementary Military Cryptography*, and *Advanced Military Cryptography*. These texts deal exclusively with *cryptography* as defined therein; that is, with the various types of ciphers and codes, their principles of construction, and their employment in cryptographing and decryptographing messages. Particular emphasis was placed upon such means and methods as are practicable for military usage. It is also assumed that the student has firmly in mind the technically precise, special nomenclature employed in those texts, for the terms and definitions therein will all be used in the present text, with essentially the same significances. If this is not the case, it is recommended that the student review his preceding work, in order to regain a familiarity with the specific meanings assigned to the terms used therein. There will be no opportunity herein to repeat this information and unless he understands clearly the significance of the terms employed, his progress will be retarded.

b. This text constitutes the first of a series of texts on *cryptanalysis*. Although most of the information contained herein is applicable to cryptograms of various types and sources, special emphasis will be laid upon the principles and methods of solving military cryptograms. Except for an introductory discussion of fundamental principles underlying the science of cryptanalytics, this first text in the series will deal solely with the principles and methods for the analysis of monoalphabetic substitution ciphers. Even with this limitation it will be possible to discuss only a few of the many variations of this one type; but with a firm grasp upon the general principles no difficulties should be experienced with any variations that may be encountered.

c. This and some of the succeeding texts will deal only with elementary types of cipher systems not because they may be encountered in military operations but because their study is essential to an understanding of the principles underlying the solution of the modern, very much more complex types of ciphers and codes that are likely to be employed by the larger governments today in the conduct of their military affairs in time of war.

d. All of this series of texts will deal only with the solution of visible secret writing. At some future date, texts dealing with the solution of invisible secret writing, and with secret signalling systems, may be prepared.

2. Mental equipment necessary for cryptanalytic work.—*a.* Captain Parker Hitt, in the first United States Army manual [1] dealing with cryptography, opens the first chapter of his valuable treatise with the following sentence:

> Success in dealing with unknown ciphers is measured by these four things in the order named: perseverance, careful methods of analysis, intuition, luck.

[1] Hitt, Capt. Parker, *Manual for the Solution of Military Ciphers*. Army Service Schools Press, Fort Leavenworth, Kansas, 1916. 2d Edition, 1918. (Both out of print.)

These words are as true today as they were then. There is no royal road to success in the solution of cryptograms. Hitt goes on to say:

> Cipher work will have little permanent attraction for one who expects results at once, without labor, for there is a vast amount of purely routine labor in the preparation of frequency tables, the rearrangement of ciphers for examination, and the trial and fitting of letter to letter before the message begins to appear.

The present author deems it advisable to add that the kind of work involved in solving cryptograms is not at all similar to that involved in solving "cross-word puzzles", for example. The wide vogue the latter have had and continue to have is due to the appeal they make to the quite common instinct for mysteries of one sort or another; but in solving a cross-word puzzle there is usually no necessity for performing any preliminary labor, and palpable results become evident after the first minute or two of attention. This successful start spurs the cross-word "addict" on to complete the solution, which rarely requires more than an hour's time. Furthermore, cross-word puzzles are all alike in basic principle and once understood, there is no more to learn. Skill comes largely from the embellishment of one's vocabulary, though, to be sure, constant practice and exercise of the imagination contribute to the ease and rapidity with which solutions are generally reached. In solving cryptograms, however, many principles must be learned, for there are many different systems of varying degrees of complexity. Even some of the simpler varieties require the preparation of tabulations of one sort or another, which many people find irksome; moreover, it is only toward the very close of the solution that results in the form of intelligible text become evident. Often, indeed, the student will not even known whether he is on the right track until he has performed a large amount of preliminary "spade work" involving many hours of labor. Thus, without at least a willingness to pursue a fair amount of theoretical study, and a *more than average amount of patience and perseverance*, little skill and experience can be gained in the rather difficult art of cryptanalysis. General Givierge, the author of an excellent treatise on cryptanalysis, remarks in this connection: [2]

> The cryptanalyst's attitude must be that of William the Silent: No need to hope in order to undertake, nor to succeed in order to persevere.

b. As regards Hitt's reference to careful methods of analysis, before one can be said to be a cryptanalyst worthy of the name it is necessary that one should have firstly a sound knowledge of the basic principles of cryptanalysis, and secondly, a long, varied, and active *practical* experience in the successful application of those principles. It is not sufficient to have read treatises on this subject. One month's actual practice in solution is worth a whole year's mere reading of theoretical principles. An exceedingly important element of success in solving the more intricate ciphers is the possession of the rather unusual mental faculty designated in general terms as the power of inductive and deductive reasoning. Probably this is an inherited rather than an acquired faculty; the best sort of training for its emergence, if latent in the individual, and for its development is the study of the natural sciences, such as chemistry, physics, biology, geology, and the like. Other sciences such as linguistics and philology are also excellent. Aptitude in mathematics is quite important, more especially in the solution of ciphers than of codes.

c. An active imagination, or perhaps what Hitt and other writers call *intuition*, is essential, but mere imagination uncontrolled by a judicious spirit will more often be a hindrance than a help. In practical cryptanalysis the imaginative or intuitive faculties must, in other words, be guided by good judgment, by practical experience, and by as thorough a knowledge of the general situation or extraneous circumstances that led to the sending of the cryptogram as is possible to obtain. In this respect the many cryptograms exchanged between correspondents whose identities and general affairs, commercial, social, or political, are known are far more readily

[2] Givierge, Général Marcel, *Cours de Cryptographie*, Paris, 1925, p. 301.

solved than are isolated cryptograms exchanged between unknown correspondents, dealing with unknown subjects. It is obvious that in the former case there are good data upon which the intuitive powers of the cryptamalyst can be brought to bear, whereas in the latter case no such data are available. Consequently, in the absence of such data, no matter how good the imagination and intuition of the cryptanalyst, these powers are of no particular service to him. Some writers, however, regard the intuitive spirit as valuable from still another viewpoint, as may be noted in the following: [3]

> Intuition, like a flash of lightning, lasts only for a second. It generally comes when one is tormented by a difficult decipherment and when one reviews in his mind the fruitless experiments already tried. Suddenly the light breaks through and one finds after a few minutes what previous days of labor were unable to reveal.

This, too, is true, but unfortunately there is no way in which the intuition may be summoned at will, when it is most needed.[4] There are certain authors who regard as indispensable the possession of a somewhat rare, rather mysterious faculty that they designate by the word "flair", or by the expression "cipher brains." Even so excellent an authority as General Givierge,[5] in referring to this mental facility, uses the following words: "Over and above perseverance and this aptitude of mind which some authors consider a special gift, and which they call intuition, or even, in its highest manifestation, clairvoyance, cryptographic studies will continue more and more to demand the qualities of orderliness and memory." Although the present author believes a special aptitude for the work is essential to cryptanalytic success, he is sure there is nothing mysterious about the matter at all. Special aptitude is prerequisite to success in all fields of endeavor. There are, for example, thousands of physicists, hundreds of excellent ones, but only a handful of world-wide fame. Should it be said, then, that a physicist

---

[3] Lange et Soudart, *Traité de Cryptographie*, Librairie Félix Alcan, Paris, 1925, p. 104.

[4] The following extracts are of interest in this connection:

The fact that the scientific investigator works 50 per cent of his time by non-rational means is, it seems, quite insufficiently recognized. There is without the least doubt an instinct for research, and often the most successful investigators of nature are quite unable to give an account of their reasons for doing such and such an experiment, or for placing side by side two apparently unrelated facts. Again, one of the most salient traits in the character of the successful scientific worker is the capacity for knowing that a point is proved when it would not appear to be proved to an outside intelligence functioning in a purely rational manner; thus the investigator feels that some proposition is true, and proceeds at once to the next set of experiments without waiting and wasting time in the elaboration of the formal proof of the point which heavier minds would need. Questionless such a scientific intuition may and does sometimes lead investigators astray, but it is quite certain that if they did not widely make use of it, they would not get a quarter as far as they do. Experiments confirm each other, and a false step is usually soon discovered. And not only by this partial replacement of reason by intuition does the work of science go on, but also to the born scientific worker—and emphatically they cannot be made—the structure of the method of research is as it were given, he cannot explain it to you, though he may be brought to agree *a postiori* to a formal logical presentation of the way the method works.—Excerpt from Needham, Joseph, *The Sceptical Biologist*, London, 1929, p. 79.

The essence of scientific method, quite simply, is to try to see how data arrange themselves into causal configurations. Scientific problems are solved by collecting data and by "thinking about them all the time." We need to look at strange things until, by the appearance of known configurations, they seem familiar, and to look at familiar things until we see novel configurations which make them appear strange. We must look at events until they become luminous. That is scientific method . . . Insight is the touchstone . . . The application of insight as the touchstone of method enables us to evaluate properly the role of imagination in scientific method. The scientific process is akin to the artistic process: it is a process of selecting out those elements of experience which fit together and recombining them in the mind. Much of this kind of research is simply a ceaseless mulling over, and even the physical scientist has considerable need of an armchair . . . Our view of scientific method as a struggle to obtain insight forces the admission that science is half art . . . Insight is the unknown quantity which has eluded students of scientific method.—Excerpts from an article entitled *Insight and Scientific Method*, by Willard Waller, in *The American Journal of Sociology*, Vol. XL, 1934.

[5] *Op. cit.*, p. 302.

who has achieved very notable success in his field has done so because he is the fortunate posessor of a *mysterious* faculty? That he is fortunate in possessing a special aptitude for his subject is granted, but that there is anything mysterious about it, partaking of the nature of clairvoyance (if, indeed, the latter is a *reality*) is not granted. While the ultimate nature of any mental process seems to be as complete a mystery today as it has ever been, the present author would like to see the superficial veil of mystery removed from a subject that has been shrouded in mystery from even before the Middle Ages down to our own times. (The principal and easily understandable reason for this is that governments have always closely guarded cryptographic secrets and anything so guarded soon becomes "mysterious.") He would, rather, have the student approach the subject as he might approach any other science that can stand on its own merits with other sciences, because cryptanalytics, like other sciences, has a practical importance in human affairs. It presents to the inquiring mind an interest in its own right as a branch of knowledge; it, too, holds forth many difficulties and disappointments, and these are all the more keenly felt when the nature of these difficulties is not understood by those unfamiliar with the special circumstances that very often are the real factors that led to success in other cases. Finally, just as in the other sciences wherein many men labor long and earnestly for the true satisfaction and pleasure that comes from work well-done, so the mental pleasure that the successful cryptanalyst derives from his accomplishments is very often the only reward for much of the drudgery that he must do in his daily work. General Givierge's words in this connection are well worth quoting:[6]

> Some studies will last for years before bearing fruit. In the case of others, cryptanalysts undertaking them never get any result. But, for a cryptanalyst who likes the work, the joy of discoveries effaces the memory of his hours of doubt and impatience.

d. With his usual deft touch, Hitt says of the element of luck, as regards the role it plays in analysis:

> As to luck, there is the old miners' proverb: "Gold is where you find it."

The cryptanalyst is lucky when one of the correspondents whose ciphers he is studying makes a blunder that gives the necessary clue; or when he finds two cryptograms identical in text but in different keys in the same system; or when he finds two cryptograms identical in text but in different systems, and so on. The element of luck is there, to be sure, *but the cryptanalyst must be on the alert* if he is to profit by these lucky "breaks."

e. If the present author were asked to state, in view of the progress in the field since 1916, what elements might be added to the four ingredients Hitt thought essential to cryptanalytic success, he would be inclined to mention the following:

(1) A broad, general education, embodying interests covering as many fields of practical knowledge as possible. This is useful because the cryptanalyst is often called upon to solve messages dealing with the most varied of human activities, and the more he knows about these activities, the easier his task.

(2) Access to a large library of current literature, and wide and direct contacts with sources of collateral information. These often afford clues as to the contents of specific messages. For example, to be able instantly to have at his disposal a newspaper report or a personal report of events described or referred to in a message under investigation goes a long way toward simplifying or facilitating solution. Government cryptanalysts are sometimes fortunately situated in this respect, especially where various agencies work in harmony.

(3) Proper coordination of effort. This includes the organization of cryptanalytic personnel into harmonious, efficient teams of cooperating individuals.

---

[6] *Op. cit.*, p. 301.

(4) Under mental equipment he would also include the faculty of being able to concentrate on a problem for rather long periods of time, without distraction, nervous irritability, and impatience. The strain under which cryptanalytic studies are necessarily conducted is quite severe and too long-continued application has the effect of draining nervous energy to an unwholesome degree, so that a word or two of caution may not here be out of place. One should continue at work only so long as a peaceful, calm spirit prevails, whether the work is fruitful or not. But just as soon as the mind becomes wearied with the exertion, or just as soon as a feeling of hopelessness or mental fatigue intervenes, it is better to stop completely and turn to other activities, rest, or play. It is essential to remark that systematization and orderliness of work are aids in reducing nervous tension and irritability. On this account it is better to take the time to prepare the data carefully, rewrite the text if necessary, and so on, rather than work with slipshod, incomplete, or improperly arranged material.

(5) A retentive memory is an important asset to cryptanalytic skill, especially in the solution of codes. The ability to remember individual groups, their approximate locations in other messages, the associations they form with other groups, their peculiarities and similarities, saves much wear and tear of the mental machinery, as well as much time in looking up these groups in indexes.

f. It may be advisable to add a word or two at this point to prepare the student to expect slight mental jars and tensions which will almost inevitably come to him in the conscientious study of this and the subsequent texts. The present author is well aware of the complaint of students that authors of texts on cryptanalysis base much of their explanation upon their foreknowledge of the "answer"—which the student does not know while he is attempting to follow the solution with an unbiased mind. They complain too that these authors use such expressions as "obviously", "naturally", "of course", "it is evident that", and so on, when the circumstances seem not at all to warrant their use. There is no question but that this sort of treatment is apt to discourage the student, especially when the point elucidated becomes clear to *him* only after many hours' labor, whereas, according to the book, the author noted the weak spot at the first moment's inspection. The present author can only promise to try to avoid making the steps appear to be much more simple than they really are, and to suppress glaring instances of unjustifiable "jumping at conclusions." At the same time he must indicate that for pedagogical reasons in many cases a message has been consciously "manipulated" so as to allow certain principles to become more obvious in the illustrative examples than they ever are in practical work. During the course of some of the explanations attention will even be directed to cases of unjustified inferences. Furthermore, of the student who is quick in observation and deduction, the author will only ask that he bear in mind that if the elucidation of certain principles seems prolix and occupies more space than necessary, this is occasioned by the author's desire to carry the explanation forward in very short, easily-comprehended, and plainly-described steps, for the benefit of students who are perhaps a bit slower to grasp but who, once they understand, are able to retain and apply principles slowly learned just as well, if not better than the students who learn more quickly.

3. Validity of results of cryptanalysis.—Valid, or authentic cryptanalytic solutions cannot and do not represent "opinions" of the cryptanalyst. They are valid only so far as they are wholly objective, and are susceptible of demonstration and proof, employing authentic, objective methods. It should hardly be necessary (but an attitude frequently encountered among laymen makes it advisable) to indicate that the validity of the results achieved by any serious cryptanalytic studies on authentic material rests upon the same sure foundations and are reached by the same general steps as the results achieved by any other scientific studies; viz, observation, hypothesis, deduction and induction, and confirmatory experiment. Implied in the latter is the

possibility that two or more qualified investigators, each working independently upon the same material, will achieve identical (or practically identical) results. Occasionally a pseudo-cryptanalyst offers "solutions" which cannot withstand such tests; a second, unbiased, investigator working independently either cannot *consistently* apply the methods alleged to have been applied by the pseudo-cryptanalyst, or else, if he can apply them at all, the results (plain-text translations) are far different in the two cases. The reason for this is that in such cases it is generally found that the "methods" are not clear-cut, straightforward or mathematical in character. Instead, they often involve the making of judgments on matters too tenuous to measure, weigh, or otherwise subject to careful scrutiny. In such cases, the conclusion to which the unprejudiced observer is forced to come is that the alleged "solution" obtained by the first investigator, the pseudo-cryptanalyst, is purely subjective. In nearly all cases where this has happened (and they occur from time to time) there has been uncovered nothing which can in any way be used to impugn the integrity of the pseudo-cryptanalyst. The worst that can be said of him is that he has become a victim of a special or peculiar form of self-delusion, and that his desire to solve the problem, usually in accord with some previously-formed opinion, or notion, has over-balanced, or undermined, his judgment and good sense.[7]

---

[7] Specific reference can be made to the following typical "case histories":

Donnelly, Ignatius, *The Great Cryptogram.* Chicago, 1888.

Owen, Orville W., *Sir Francis Bacon's Cipher Story.* Detroit, 1895.

Gallup, Elizabeth Wells, *Francis Bacon's Bilateral Cipher.* Detroit, 1900.

Margoliouth, D. S., *The Homer of Aristotle.* Oxford, 1923.

Newbold, William Romaine, *The Cipher of Roger Bacon.* Philadelphia, 1928. (For a scholarly and complete demolition of Professor Newbold's work, see an article entitled *Roger Bacon and the Voynich MS*, by John M. Manly, in *Speculum*, Vol. *VI*, No. 3, July 1931.)

Arensberg, Walter Conrad, *The Cryptography of Shakespeare.* Los Angeles, 1922.

The Shakespearean Mystery. Pittsburgh, 1928.

The Baconian Keys. Pittsburgh, 1928.

Feely, Joseph Martin, *The Shakespearean Cypher.* Rochester, N. Y., 1931.

Deciphering Shakespeare. Rochester, N. Y., 1934.

SECTION II

FUNDAMENTAL PRINCIPLES

	Paragraph
The four basic operations in cryptanalysis	4
The determination of the language employed	5
The determination of the general system	6
The reconstruction of the specific key	7
The reconstruction of the plain text	8

4. The four basic operations in cryptanalysis.—*a.* The solution of practically every cryptogram involves four fundamental operations or steps:

(1) The determination of the language employed in the plain-text version.

(2) The determination of the general system of cryptography employed.

(3) The reconstruction of the specific key in the case of a cipher system, or the reconstruction, partial or complete, of the code book, in the case of a code system; or both, in the case of an enciphered code system.

(4) The reconstruction or establishment of the plain text.

b. These operations will be taken up in the order in which they are given above and in which they usually are performed in the solution of cryptograms, although occasionally the second step may precede the first.

5. The determination of the language employed.—*a.* There is not much that need be said with respect to this operation except that the determination of the language employed seldom comes into question in the case of studies made of the cryptograms of an organized enemy. By this is meant that during wartime the enemy is of course known, and it follows, therefore, that the language he employs in his messages will almost certainly be his native or mother tongue. Only occasionally nowadays is this rule broken. Formerly it often happened, or it might have indeed been the general rule, that the language used in diplomatic correspondence was not the mother tongue, but French. In isolated instances during the World War, the Germans used English when their own language could for one reason or another not be employed. For example, for a year or two before the entry of the United States into that war, during the time America was neutral and the German Government maintained its embassy in Washington, some of the messages exchanged between the Foreign Office in Berlin and the Embassy in Washington were cryptographed in English, and a copy of the code used was deposited with the Department of State and our censor. Another instance is found in the case of certain Hindu conspirators who were associated with and partially financed by the German Government in 1915 and 1916; they employed English as the language of their cryptographic messages. Occasionally the cryptograms of enemy agents may be in a language different from that of the enemy. But in general these are, as has been said, isolated instances; as a rule, the language used in cryptograms exchanged between members of large organizations is the mother tongue of the correspondents. Where this is not the case, that is, when cryptograms of unknown origin must be studied, the cryptanalyst looks for any indications on the cryptograms themselves which may lead to a conclusion as to the language employed. Address, signature, and plain-language words in the preamble or in the body of the text all come under careful scrutiny, as well as all extraneous circumstances connected with the manner in which the cryptograms were obtained, the person on whom they were found, or the locale of their origin and destination.

b. In special cases, or under special circumstances a clue to the language employed is found in the nature and composition of the cryptographic text itself. For example, if the letters K and W are entirely absent or appear very rarely in messages, it may indicate that the language is Spanish, for these letters are absent in the alphabet of that language and are used only to spell foreign words or names. The presence of accented letters or letters marked with special signs of one sort or another, peculiar to certain languages, will sometimes indicate the language used. The Japanese Morse telegraph alphabet and the Russian Morse telegraph alphabet contain combinations of dots and dashes which are peculiar to those alphabets and thus the interception of messages containing these special Morse combinations at once indicates the language involved. Finally, there are certain peculiarities of alphabetic languages which, in certain types of cryptograms, *viz*, pure transposition, give clues as to the language used. For example, the frequent digraph C H, in German, leads to the presence, in cryptograms of the type mentioned, of many isolated C's and H's; if this is noted, the cryptogram may be assumed to be in German.

c. In some cases it is perfectly possible to perform certain steps in cryptanalysis *before* the language of the cryptogram has been definitely determined. Frequency studies, for example, may be made and analytic processes performed without this knowledge, and by a cryptanalyst wholly unfamiliar with the language even if it has been identified, or who knows only enough about the language to enable him to recognize valid combinations of letters, syllables, or a few common words in that language. He may, after this, call to his assistance a translator who may not be a cryptanalyst but who can materially aid in making necessary assumptions based upon his special knowledge of the characteristics of the language in question. Thus, cooperation between cryptanalyst and translator results in solution.[1]

6. The determination of the general system.—*a.* Except in the case of the more simple types of cryptograms, the determination of the general system according to which a given cryptogram has been produced is usually a difficult, if not the most difficult, step in its solution. The reason for this is not hard to find.

b. As will become apparent to the student as he proceeds with his study, *in the final analysis, the solution of every cryptogram involving a form of substitution depends upon its reduction to monoalphabetic terms, if it is not originally in those terms.* This is true not only of ordinary substitution ciphers, but also of combined substitution-transposition ciphers, and of enciphered code. If the cryptogram must be reduced to monoalphabetic terms, the manner of its accomplishment is usually indicated by the cryptogram itself, by external or internal phenomena which become apparent to the cryptanalyst as he studies the cryptogram. If this is impossible, or too difficult the cryptanalyst must, by one means or another, discover how to accomplish this reduction, by bringing to bear all the special or collateral information he can get from all the sources at his command. If both these possibilities fail him, there is little left but the long, tedious, and often fruitless process of elimination. In the case of transposition ciphers of the more complex type, the discovery of the basic method is often simply a matter of long and tedious elimination of possibilities. For cryptanalysis has unfortunately not yet attained, and may indeed never attain, the precision found today in qualitative analysis in chemistry, for example, where the analytic process is absolutely clear cut and exact in its dichotomy. A few words in explanation of what is meant may not be amiss. When a chemist seeks to determine the identity of an unknown

[1] The writer has seen in print statements that "during the World War . . . decoded messages in Japanese and Russian without knowing a word of either language." The extent to which such statements are exaggerated will soon become obvious to the student. Of course, there are occasional instances in which a mere clerk with quite limited experience may be able to "solve" a message in an extremely simple system in a language of which he has no knowledge at all; but such a "solution" calls for nothing more arduous than the ability to recognize pronounceable combinations of vowels and consonants—an ability that hardly deserves to be rated as "cryptanalytic" in any real sense. To say that it is possible to solve a cryptogram in a foreign language "without knowing a word of that language" is not quite the same as to say that it is possible to do so with only a slight knowledge of the language; and it may be stated without cavil that the better the cryptanalyst's knowledge of the language, the greater are the chances for his success and, in any case, the easier is his work.

substance, he applies certain specific reagents to the substance and in a specific sequence. The first reagent tells him definitely into which of two primary classes the unknown substance falls. He then applies a second test with another specific reagent, which tells him again quite definitely into which of two secondary classes the unknown substance falls, and so on, until finally he has reduced the unknown substance to its simplest terms and has found out what it is. In striking contrast to this situation, cryptanalysis affords exceedingly few "reagents" or tests that may be applied to determine positively that a given cipher belongs to one or the other of two systems yielding externally similar results. And this is what makes the analysis of an isolated, complex cryptogram so difficult. Note the limiting adjective "isolated" in the foregoing sentence, for it is used advisedly. It is not often that the general system fails to disclose itself or cannot be discovered by painstaking investigation when there is a great volume of text accumulating from a regular traffic between numerous correspondents in a large organization. *Sooner or later* the system becomes known, either because of blunders and carelessness on the part of the personnel entrusted with the cryptographing of the messages, or because the accumulation of text itself makes possible the determination of the general system by cryptanalytic studies. But in the case of a single or even a few isolated cryptograms concerning which little or no information can be gained by the cryptanalyst, he is often unable, without a knowledge of, or a shrewd guess as to the general system employed, to decompose the heterogeneous text of the cryptogram into homogeneous, monoalphabetic text, which is the ultimate and essential step in analysis. The only knowledge that the cryptanalyst can bring to his aid in this most difficult step is that gained by long experience and practice in the analysis of many different types of systems.

c. On account of the complexities surrounding this particular phase of cryptanalysis, and because in any scheme of analysis based upon successive eliminations of alternatives the cryptanalyst can only progress so far as the extent of his own knowledge of *all* the possible alternatives will permit, it is necessary that detailed discussion of the eliminative process be postponed until the student has covered most of the field. For example, the student will perhaps want to know at once how he can distinguish between a cryptogram that is in code or enciphered code from one that is in cipher. It is at this stage of his studies impracticable to give him any helpful indications on his question. In return it may be asked of him why he should expect to be able to do this in the early stages of his studies when often the experienced expert cryptanalyst is baffled on the same score!

d. Nevertheless, in lieu of more precise tests not yet discovered, a general guide that may be useful in cryptanalysis will be built up, step by step as the student progresses, in the form of a series of charts comprising what may be designated *An Analytical Key For Cryptanalysis.* (See Par. 50.) It may be of assistance to the student if, as he proceeds, he will carefully study the charts and note the place which the particular cipher he is solving occupies in the general cryptanalytic panorama. These charts admittedly constitute only very brief outlines, and can therefore be of but little direct assistance to him in the analysis of the more complex types of ciphers he may encounter later on. So far as they go, however, they may be found to be quite useful in the study of elementary cryptanalysis. For the experienced cryptanalyst they can serve only as a means of assuring that no possible step or process is inadvertently overlooked in attempts to solve a difficult cipher.

e. Much of the labor involved in cryptanalytic work, as referred to in Par. 2, is connected with this determination of the general system. The preparation of the text, its rewriting in different forms, sometimes being rewritten in a half dozen ways, the recording of letters, the establishment of frequencies of occurrences of letters, comparisons and experiments made with known material of similar character, and so on, constitute much labor that is most often indispensable, but which sometimes turns out to have been wholly unnecessary, or in vain. In a

recent treatise [2] it is stated quite boldly that "this work once done, the determination of the system is often relatively easy." This statement can certainly apply only to the simpler types of ciphers; it is entirely misleading as regards the much more frequently encountered complex cryptograms of modern times.

7. The reconstruction of the specific key.—*a.* Nearly all practical cryptographic methods require the use of a specific key to guide, control, or modify the various steps under the general system. Once the latter has been disclosed, discovered, or has otherwise come into the possession of the cryptanalyst, the next step in solution is to determine, if necessary, and if possible, the specific key that was employed to cryptograph the message or messages under examination. This determination may not be in complete detail; it may go only so far as to lead to a knowledge of the number of alphabets involved in a substitution cipher, or the number of columns involved in a transposition cipher, or that a one-part code has been used, in the case of a code system. But it is often desirable to determine the specific key in as complete a form and with as much detail as possible, for this information will very frequently be useful in the solution of subsequent cryptograms exchanged between the same correspondents, since the nature of the specific key in a solved case may be expected to give clues to the specific key in an unsolved case.

b. Frequently, however, the reconstruction of the key is not a prerequisite to, and does not constitute an absolutely necessary preliminary step in, the fourth basic operation, *viz*, the reconstruction or establishment of the plain text. In many cases, indeed, the two processes are carried along simultaneously, the one assisting the other, until in the final stages both have been completed in their entireties. In still other cases the reconstruction of the specific key may succeed instead of precede the reconstruction of the plain text, and is accomplished purely as a matter of academic interest; or the specific key may, in unusual cases, never be reconstructed.

8. The reconstruction of the plain text.—*a.* Little need be said at this point on this phase of cryptanalysis. The process usually consists, in the case of substitution ciphers, in the establishment of equivalency between specific letters of the cipher text and the plain text, letter by letter, pair by pair, and so on, depending upon the particular type of substitution system involved. In the case of transposition ciphers, the process consists in rearranging the elements of the cipher text, letter by letter, pair by pair, or occasionally word by word, depending upon the particular type of transposition system involved, until the letters or words have been returned to their original plain-text order. In the case of code, the process consists in determining the meaning of each code group and inserting this meaning in the code text to reestablish the original plain text.

b. The foregoing processes do not, as a rule, begin at the beginning of a message and continue letter by letter, or group by group in sequence up to the very end of the message. The establishment of values of cipher letters in substitution methods, or of the positions to which cipher letters should be transferred to form the plain text in the case of transposition methods, comes at very irregular intervals in the process. At first only one or two values scattered here and there throughout the text may appear; these then form the "skeletons" of words, upon which further work, by a continuation of the reconstruction process, is made possible; in the end the complete or nearly complete [3] text is established.

c. In the case of cryptograms in a foreign language, the translation of the solved messages is a final and necessary step, but is not to be considered as a cryptanalytic process. However, it is commonly the case that the translation process will be carried on simultaneously with the cryptanalytic, and will aid the latter, especially when there are lacunae which may be filled in from the context. (See also Par. *5c* in this connection.)

[2] Lange et Soudart, *op. cit.*, p. 106.

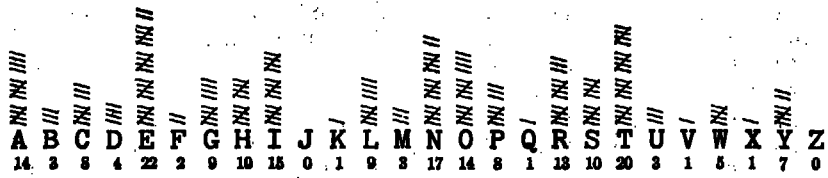
[3] Sometimes in the case of code, the meaning of a few code groups may be lacking, because there is insufficient text to establish their meaning.

SECTION III

FREQUENCY DISTRIBUTIONS

	Paragraph
The simple or uniliteral frequency distribution	9
Important features of the normal uniliteral frequency distribution	10
Constancy of the standard or normal uniliteral frequency distribution	11

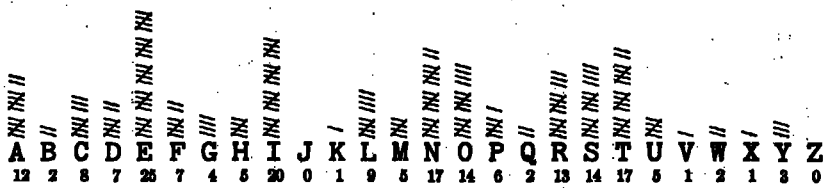
9. The simple or uniliteral frequency distribution.—*a.* It has long been known to cryptographers and typographers that the letters composing the words of any intelligible written text composed in any language which is alphabetic in construction are employed with greatly varying frequencies. For example, if on cross-section paper a simple tabulation, shown in Fig. 1, called a *uniliteral frequency distribution*, is made of the letters composing the words of the preceding sentence, the variation in frequency is strikingly demonstrated. It is seen that whereas certain letters, such as A, E, I, N, O, R, S, and T, are employed very frequently, other letters, such as C, G, P, and W are employed not nearly so frequently, while still other letters, such as F, J, Q, V, and Z are employed either seldom or not at all.



(Total=200 letters)

FIGURE 1.

b. If a similar tabulation is now made of the letters comprising the words of the second sentence in the preceding paragraph, the graph shown in Fig. 2 is obtained. Both sentences have exactly the same number of letters (200).



(Total=200 letters)

FIGURE 2.

c. Although each of these two graphs exhibits great variation in the relative frequencies with which *different* letters are employed in the respective sentences to which they apply, no marked differences are exhibited between the frequencies of the *same* letter in the two graphs. Compare, for example, the frequencies of A, B, C . . . Z in Fig. 1 with those of A, B, C, . . . Z in Fig. 2. Aside from one or two exceptions, as in the case of the letter F, these two graphs agree rather strikingly.

d. This agreement, or *similarity*, would be practically complete if the two texts were much longer, for example, five times as long. In fact, when two texts of similar character, each containing more than 1,000 letters, are compared, it would be found that the respective frequencies of the 26 letters composing the two graphs show only very slight differences. This means, in other words, that in normal text each letter of the alphabet occurs with a rather *constant* or *characteristic frequency* which it tends to approximate, depending upon the length of the text analyzed. The longer the text (within certain limits), the closer will be the approximation.[1]

e. An experiment along these lines will be convincing. A series of 260 official telegrams [2] passing through the War Department Message Center was examined statistically. The messages were divided into five sets, each totaling 10,000 letters, and the five distributions shown in Table 1–A, were obtained.

f. If the five distributions in Table 1–A are summed, the results are as shown in Table 2–A.

TABLE 1–A.—*Absolute frequencies of letters appearing in five sets of Governmental plain-text telegrams, each set containing 10,000 letters, arranged alphabetically*

Message No. 1		Message No. 2		Message No. 3		Message No. 4		Message No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1,367	E	1,294	E	1,292	E	1,270	E	1,275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	238	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10,000		10,000		10,000		10,000		10,000

[1] See footnote 5, page 16.

[2] These comprised messages from several departments in addition to the War Department and were all of an administrative character.

TABLE 2-A.—*Absolute frequencies of letters appearing in the combined five sets of messages totaling 50,000 letters, arranged alphabetically*

A	3,683	G	819	L	1,821	Q	175	V	766
B	487	H	1,694	M	1,237	R	3,788	W	780
C	1,534	I	3,676	N	3,975	S	3,058	X	231
D	2,122	J	82	O	3,764	T	4,595	Y	967
E	6,498	K	148	P	1,335	U	1,300	Z	49
F	1,416								

g. The frequencies noted in subparagraph *f*, when reduced to the basis of 1,000 letters and then used as a basis for constructing a simple chart that will exhibit the variations in frequency in a striking manner, yield the following graph which is hereafter designated as the *normal*, or *standard uniliteral frequency distribution* for English telegraphic plain text:

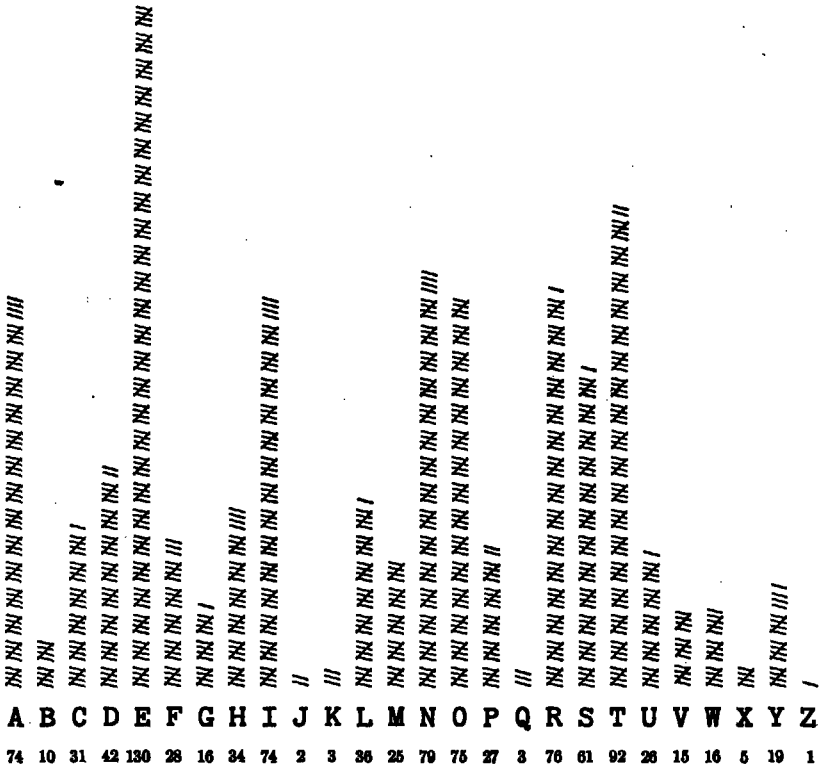


FIGURE 3.

10. Important features of the normal uniliteral frequency distribution.—*a*. When the graph shown in Fig. 3 is studied in detail, the following features are apparent:

(1) It is quite irregular in appearance. This is because the letters are used with greatly varying frequencies, as discussed in the preceding paragraph. This irregular appearance is often described by saying that the graph shows marked *crests and troughs*, that is, points of high frequency and low frequency.

148274—38——2

(2) The relative positions in which the crests and troughs fall within the graph, that is, the *spatial relations* of the crests and troughs, are rather definitely fixed and are determined by circumstances which have been explained in a preceding text.[3]

(3) The relative heights and depths of the crests and troughs within the graph, that is, the *linear extensions* of the lines marking the respective frequencies, are also rather definitely fixed, as would be found if an equal volume of similar text were analyzed.

(4) The most prominent crests are marked by the vowels A, E, I, O, and the consonants N, R, S, T; the most prominent troughs are marked by the consonants J, K, Q, X, and Z.

(5) The important data are summarized in tabular form in Table 3.

TABLE 3

	Frequency	Percent of total	Percent of total in round numbers
6 Vowels: A E I O U Y	398	39.8	40
20 Consonants:			
5 High Frequency (D N R S T)	350	35.0	35
10 Medium Frequency (B C F G H L M P V W)	238	23.8	24
5 Low Frequency (J K Q X Z)	14	1.4	1
Total	1,000	100.0	100

(6) The frequencies of the letters of the alphabet are as follows:

A	74	G	16	L	36	Q	3	V	15
B	10	H	34	M	25	R	76	W	16
C	31	I	74	N	79	S	61	X	5
D	42	J	2	O	75	T	92	Y	19
E	130	K	3	P	27	U	26	Z	1
F	28								

(7) The relative order of frequency of the letters is as follows:

E	130	I	74	C	31	Y	19	X	5
T	92	S	61	F	28	G	16	Q	3
N	79	D	42	P	27	W	16	K	3
R	76	L	36	U	26	V	15	J	2
O	75	H	34	M	25	B	10	Z	1
A	74								

(8) The four vowels A, E, I, O (combined frequency 353) and the four consonants N, R, S, T (combined frequency 308) form 661 out of every 1,000 letters of plain text; in other words, *less than ⅓ of the alphabet is employed in writing ⅔ of normal plain text.*

[3] Section VII, *Elementary Military Cryptography.*

b. The data given in Fig. 3 and Table 3 represent the relative frequencies found in a large volume of English telegraphic text of a governmental, administrative character. These frequencies will vary somewhat with the nature of the text analyzed. For example, if an equal number of telegrams dealing solely with *commercial* transactions in the *leather industry* were studied statistically, the frequencies would be slightly different because of the repeated occurrence of words peculiar to that industry. Again, if an equal number of telegrams dealing solely with *military* messages of a *tactical* character were studied statistically, the frequencies would differ slightly from those found above for general governmental messages of an administrative character.

c. If ordinary English literary text (such as may be found in any book, newspaper, or printed document) were analyzed, the frequencies of certain letters would be changed to an appreciable degree. This is because in telegraphic text words which are not strictly essential for intelligibility (such as the definite and indefinite articles, certain prepositions, conjunctions and pronouns) are omitted. In addition, certain essential words, such as "stop", "period", "comma", and the like, which are usually indicated in written or printed matter by symbols not easy to transmit telegraphically and which must, therefore, be spelled out in telegrams, occur very frequently. Furthermore, telegraphic text often employs longer and more uncommon words than does ordinary newspaper or book text.

d. As a matter of fact, other tables compiled in the Office of the Chief Signal Officer gave slightly different results, depending upon the source of the text. For example, three tables based upon 75,000, 100,000, and 136,257 letters taken from various sources (telegrams, newspapers, magazine articles, books of fiction) gave as the relative order of frequency for the first 10 letters the following:

For 75,000 letters	E T R N I O A S D L
For 100,000 letters	E T R I N O A S D L
For 136,257 letters	E T R N A O I S L D

TABLE 4.—*Frequency table for 10,000 letters of literary English, as compiled by Hitt*

ALPHABETICALLY ARRANGED

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

Hitt also compiled data for telegraphic text (but does not state what kind of messages) and gives the following table:

TABLE 5.—*Frequency table for 10,000 letters of telegraphic English, as compiled by Hitt*

ALPHABETICALLY ARRANGED

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

e. Frequency data applicable purely to English military text were compiled by Hitt,[4] from a study of 10,000 letters taken from orders and reports. The frequencies found by him are given in Tables 4 and 5.

11. Constancy of the standard or normal, uniliteral frequency distribution.—*a.* The relative frequencies disclosed by the statistical study of large volumes of text may be considered to be the standard or *normal* frequencies of the letters of written English. Counts made of smaller volumes of text will tend to approximate these normal frequencies, and, within certain limits,[5] the smaller the volume, the lower will be the degree of approximation to the normal, until, in the case of a very short message, the normal proportions may not obtain at all. It is advisable that the student fix this fact firmly in mind, for the sooner he realizes the true nature of any data relative to the frequency of occurrence of letters in text, the less often will his labors toward the solution of specific ciphers be thwarted and retarded by too strict an adherence to these generalized principles of frequency. He should constantly bear in mind that such data are merely statistical generalizations, that they will be found to hold strictly true only in large volumes of text, and that they may not even be approximated in short messages.

b. Nevertheless the normal frequency distribution or the "normal expectancy" for any alphabetic language is, in the last analysis, the best guide to, and the usual basis for, the solution of cryptograms of a certain type. It is useful, therefore, to reduce the normal, uniliteral frequency distribution to a basis that more or less closely approximates the volume of text which the cryptanalyst most often encounters in individual cryptograms. As regards length of messages, counting only the letters in the body, and excluding address and signature, a study of the

[4] *Op. cit.*, pp. 6–7.

[5] It is useless to go beyond a certain limit in establishing the normal-frequency distribution for a given language. As a striking instance of this fact, witness the frequency study made by an indefatigable German, Kaeding, who in 1898 made a count of the letters in about 11,000,000 words, totaling about 62,000,000 letters in German text. When reduced to a percentage basis, and when the relative order of frequency was determined, the results he obtained differed very little from the results obtained by Kasiski, a German cryptographer, from a count of only 1,060 letters. See Kaeding, *Haeufigkeitswoerterbuch*, Steglitz, 1898; Kasiski, *Die Geheimschriften und die Dechiffrir-Kunst*, Berlin, 1863.

260 telegrams referred to in paragraph 9 shows that the arithmetical average is 217 letters; the statistical mean, or weighted average,[6] however, is 191 letters. These two results are, however, close enough together to warrant the statement that the *average* length of telegrams is approximately 200 letters. The frequencies given in Par. 9*f* have therefore been reduced to a basis of 200 letters, and the following uniliteral frequency distribution may be taken as showing the most typical distribution to be expected in 200 letters of telegraphic English text:

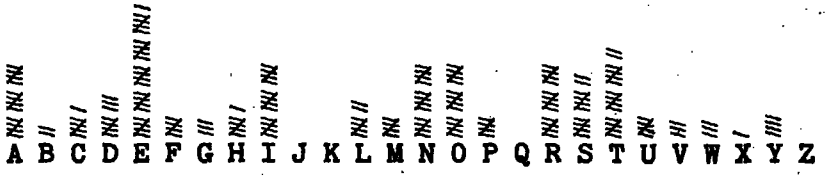


FIGURE 4.

c. The student should take careful note of the appearance of the distribution[7] shown in Fig. 4, for it will be of much assistance to him in the early stages of his study. The manner of setting down the tallies should be followed by him in making his own distributions, indicating every fifth occurrence of a letter by an oblique tally. This procedure almost automatically shows the total number of occurrences for each letter, and yet does not destroy the graphical appearance of the distribution, especially if care is taken to use approximately the same amount of space for each set of five tallies. Cross-section paper is very useful for this purpose.

d. The word "uniliteral" in the designation "uniliteral frequency distribution" means "single letter", and it is to be inferred that other types of frequency distributions may be encountered. For example, a distribution of pairs of letters, constituting a biliteral frequency distribution, is very often used in the study of certain cryptograms in which it is desired that pairs made by combining successive letters be listed. A biliteral distribution of A B C D E F would take these pairs: AB, BC, CD, DE, EF. The distribution could be made in the form of a large square divided up into 676 cells. When distributions beyond biliteral are required (triliteral, quadraliteral, etc.) they can only be made by listing them in some order, for example, alphabetically based on the 1st, 2d, 3d, . . . letter.

[6] The arithmetical average is obtained by adding each different length and dividing by the number of different-length messages; the mean is obtained by multiplying each different length by the number of messages of that length, adding all products, and dividing by the total number of messages.

[7] The use of the terms "distribution" and "frequency distribution", instead of "table" and "frequency table", respectively, is considered advisable from the point of view of consistency with the usual statistical nomenclature. When data are given in tabular form, with frequencies indicated by numbers, then they may properly be said to be set out in the form of a *table*. When, however, the same data are distributed in a chart which partakes of the nature of a graph, with the data indicated by horizontal or vertical linear extensions, or by a curve connecting points corresponding to quantities, then it is more proper to call such a graphic representation of the data a *distribution*.

SECTION IV

FUNDAMENTAL USES OF THE UNILITERAL FREQUENCY DISTRIBUTION

	Paragraph
The four facts which can be determined from a study of the uniliteral frequency distribution for a cryptogram	12
Determining the class to which a cipher belongs	13
Determining whether a substitution cipher is monoalphabetic or polyalphabetic	14
Determining whether the cipher alphabet is a standard, or a mixed cipher alphabet	15
Determining whether the standard cipher alphabet is direct or reversed	16

12. The four facts which can be determined from a study of the uniliteral frequency distribution for a cryptogram. *a.* The following four facts (to be explained subsequently) can usually be determined from an inspection of the uniliteral frequency distribution for a given cipher message of average length, composed of letters:

(1) Whether the cipher belongs to the substitution or the transposition class;

(2) If to the former, whether it is monoalphabetic or polyalphabetic in character;

(3) If monoalphabetic, whether the cipher alphabet is a standard cipher alphabet or a mixed cipher alphabet;

(4) If standard, whether it is a direct or reversed standard cipher alphabet.

b. For immediate purposes the first two of the foregoing determinations are quite important and will be discussed in detail in the next two subparagraphs; the other two determinations will be touched upon very briefly, leaving their detailed discussion for subsequent sections of the text.

13. Determining the class to which a cipher belongs.—*a.* The determination of the class to which a cipher belongs is usually a relatively easy matter because of the fundamental difference in the nature of transposition and of substitution as cryptographic processes. In a transposition cipher the original letters of the plain text have merely been rearranged, without any change whatsoever in their identities, that is, in the conventional values they have in the normal alphabet. Hence, the numbers of vowels (A, E, I, O, U, Y), high-frequency consonants (D, N, R, S, T), medium-frequency consonants (B, C, F, G, H, L, M, P, V, W), and low-frequency consonants (J, K, Q, X, Z) are exactly the same in the cryptogram as they are in the plain-text message. Therefore, the percentages of vowels, high, medium, and low-frequency consonants are the same in the transposed text as in the equivalent plain text. In a substitution cipher, on the other hand, the identities of the original letters of the plain text have been changed, that is, the conventional values they have in the normal alphabet have been altered. Consequently, if a count is made of the various letters present in such a cryptogram, it will be found that the number of vowels, high, medium, and low-frequency consonants will usually be quite different in the cryptogram from what they are in the original plain-text message. Therefore, the percentages of vowels, high, medium, and low-frequency consonants are usually quite different in the substitution text from what they are in the equivalent plain text. From these considerations it follows that if in a specific cryptogram the percentages of vowels, high, medium, and low-frequency consonants are approximately the same as would be expected in normal plain text, the cryptogram *probably* belongs to the transposition class; if these percentages are quite different from those to be expected in normal plain text the cryptogram *probably* belongs to the substitution class.

b. In the preceding subparagraph the word "probably" was emphasized by italicizing it, for there can be no certainty in every case of this determination. *Usually* these percentages in a transposition cipher are close to the normal percentages for plain text; *usually*, in a substitution cipher, they are far different from the normal percentages for plain text. But occasionally a cipher message is encountered which is difficult to classify with a reasonable degree of certainty because the message is too short for the general principles of frequency to manifest themselves. It is clear that if in actual messages there were no variation whatever from the normal vowel and consonant percentages given in Table 3, the determination of the class to which a specific cryptogram belongs would be an extremely simple matter. But unfortunately there is always some variation or deviation from the normal. Intuition suggests that as messages decrease in length there may be a greater and greater departure from the normal proportions of vowels, high, medium, and low-frequency consonants, until in very short messages the normal proportions may not hold at all. Similarly, as messages increase in length there may be a lesser and lesser departure from the normal proportions, until in messages totalling a thousand or more letters there may be no difference at all between the actual and the theoretical proportions. But intuition is not enough, for in dealing with specific messages of the length of those commonly encountered in practical work the question sometimes arises as to exactly how much deviation (from the normal proportions) may be allowed for in a cryptogram which shows a considerable amount of deviation from the normal and which might still belong to the transposition rather than to the substitution class.

c. Statistical studies have been made on this matter and some graphs have been constructed thereon. These are shown in Charts 1–4 in the form of simple curves, the use of which will now be explained. Each chart contains two curves marking the lower and upper limits, respectively, of the theoretical amount of deviation (from the normal percentages) of vowels or consonants which may be allowable in a cipher believed to belong to the transposition class.

d. In Chart 1, curve V_1 marks the lower limit of the theoretical amount of deviation from the normal number of vowels to be expected in a message of given length; curve V_2 marks the upper limit of the same thing. Thus, for example, in a message of 100 letters in plain English there should be between 33 and 47 vowels (A E I O U Y). Likewise, in Chart 2 curves H_1 and H_2 mark the lower and upper limits as regards the high-frequency consonants. In a message of 100 letters there should be between 28 and 42 high-frequency consonants (D N R S T). In Chart 3, curves M_1 and M_2 mark the lower and upper limits as regards the medium-frequency consonants. In a message of 100 letters there should be between 17 and 31 medium-frequency consonants (B C F G H L M P V W). Finally, in Chart 4, curves L_1 and L_2 mark the lower and upper limits as regards the low-frequency consonants. In a message of 100 letters there should be between 0 and 3 low-frequency consonants (J K Q X Z). In using the charts, therefore, one finds the point of intersection of the vertical coordinate corresponding to the length of the message, with the horizontal coordinate corresponding to (1) the number of vowels, (2) the number of high-frequency consonants, (3) the number of medium-frequency consonants, and (4) the number of low-frequency consonants actually counted in the message. If all four points of intersection fall within the area delimited by the respective curves, then the number of vowels, high, medium, and low-frequency consonants corresponds with the number theoretically expected in a normal plain-text message of the same length; since the message under investigation is not plain text, it follows that the cryptogram may certainly be classified as a transposition cipher. On the other hand, if one or more of these points of intersection falls outside the area delimited by the respective curves, it follows that the cryptogram is probably a substitution cipher. The distance that the point of intersection falls outside the area delimited by these curves is a more or less rough measure of the improbability of the cryptogram's being a transposition cipher.

e. Sometimes a cryptogram is encountered which is hard to classify with certainty even with the foregoing aids, because it has been consciously prepared with a view to making the classification difficult. This can be done either by selecting peculiar words (as in "trick cryptograms") or by employing a cipher alphabet in which letters of *approximately similar normal frequencies* have been interchanged. For example, E may be replaced by O, T by R, and so on, thus yielding

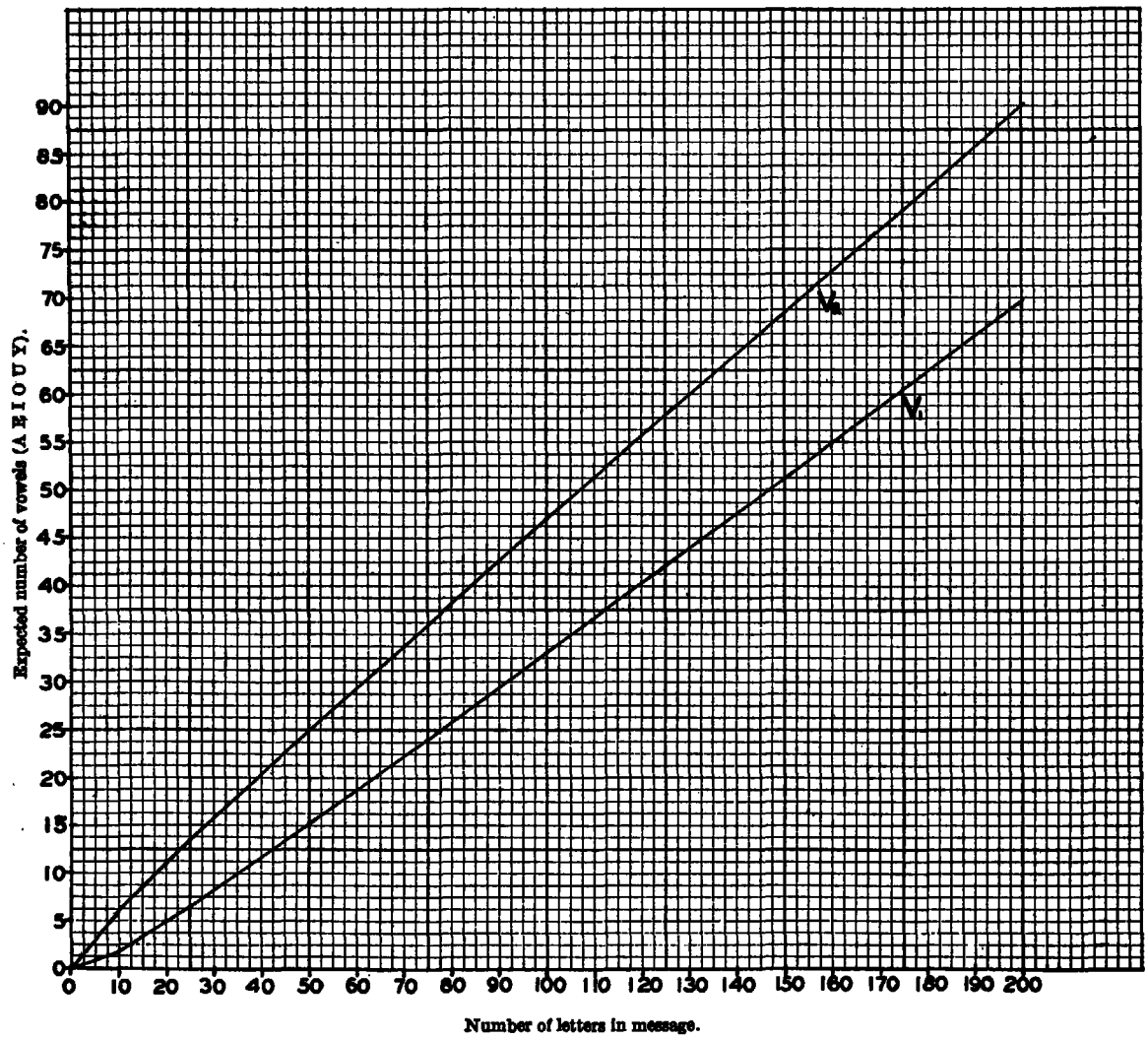


CHART No. 1.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of vowels to be expected in messages of various lengths. (See Par. 13*d*.)

a cryptogram giving external indications of being a transposition cipher but which is really a substitution cipher. If the cryptogram is not too short, a close study will usually disclose what has been done, as well as the futility of so simple a subterfuge.

f. In the majority of cases, in practical work, the determination of the class to which a cipher of average length belongs can be made from a mere inspection of the message, after the cryptanalyst has acquired a familiarity with the normal appearance of transposition and of substitution ciphers. In the former case, his eyes very speedily note many high-frequency letters, such as E, T, N, R, O, and S, with the absence of low-frequency letters, such as J, K, Q, X,

and Z; in the latter case, his eyes just as quickly note the presence of many low-frequency letters, and a corresponding absence of the usual high-frequency letters.

g. Another rather quickly completed test, in the case of the simpler varieties of ciphers, is to look for *repetitions of groups of letters*. As will become apparent very soon, recurrences of syllables, entire words and short phrases constitute a characteristic of all normal plain text. Since a transposition cipher involves a change in the *sequence* of the letters composing a plain-

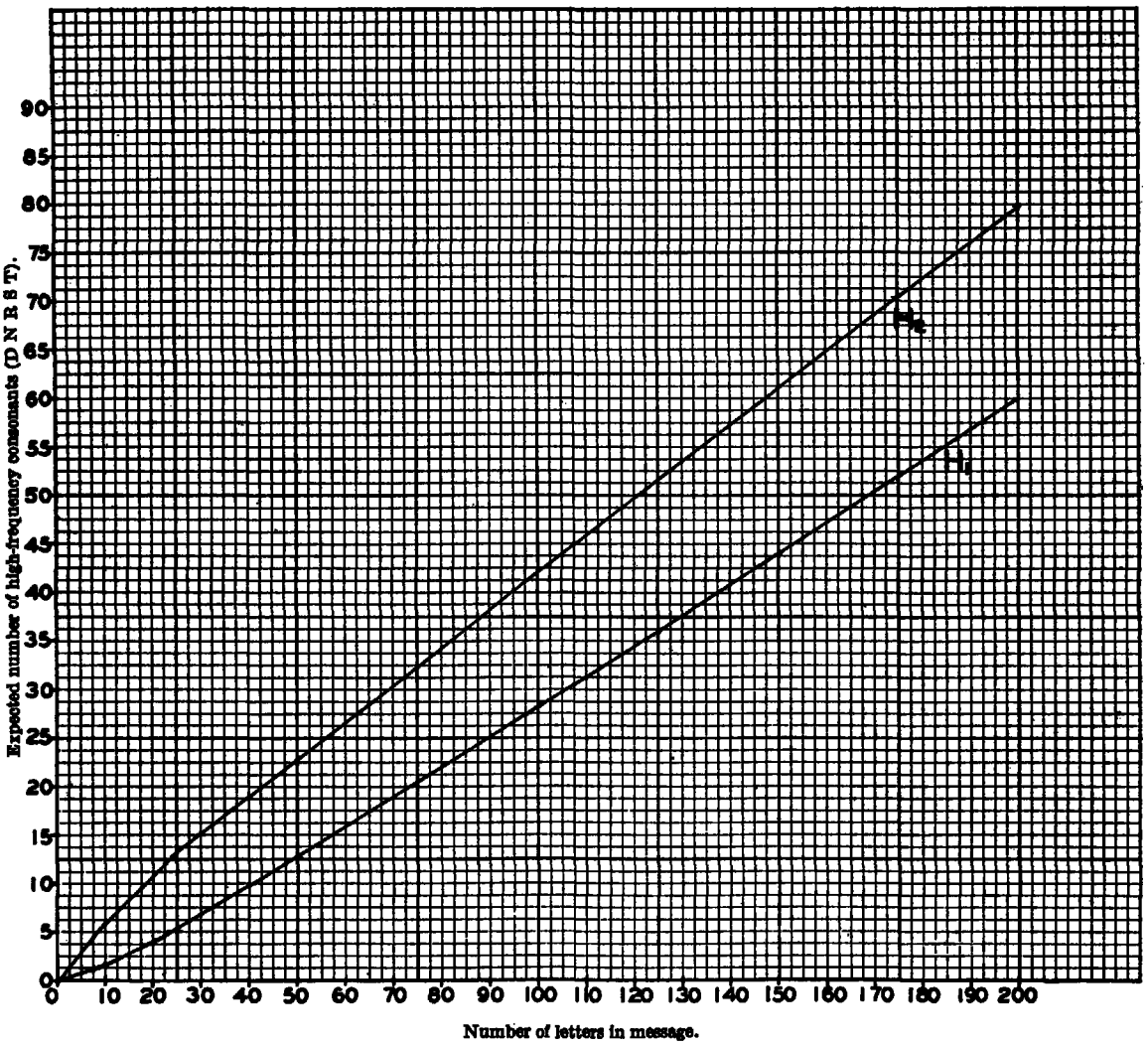


CHART No. 2.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of high-frequency conso nants to be expected in messages of various lengths. (See Par. 13d.)

text message, such recurrences are broken up so that the cipher text no longer will show repetitions of more or less lengthy sequences of letters. But if a cipher message does show many repetitions and these are of several letters in length, say over four or five, the conclusion is at once warranted that the cryptogram is most probably a substitution and not a transposition cipher. However, for the beginner in cryptanalysis, it will be advisable to make the uniliteral frequency distribution, and note the frequencies of the vowels, the high, medium, and low-frequency consonants. Then, referring to Charts 1 to 4, he should carefully note whether or not the observed frequencies for

these categories of letters fall within the limits of the theoretical frequencies for a normal plain-text message of the same length, and be guided accordingly.

h. It is obvious that the foregoing rule applies only to ciphers composed wholly of letters. If a message is composed entirely of figures, or of arbitrary signs and symbols, or of intermixtures

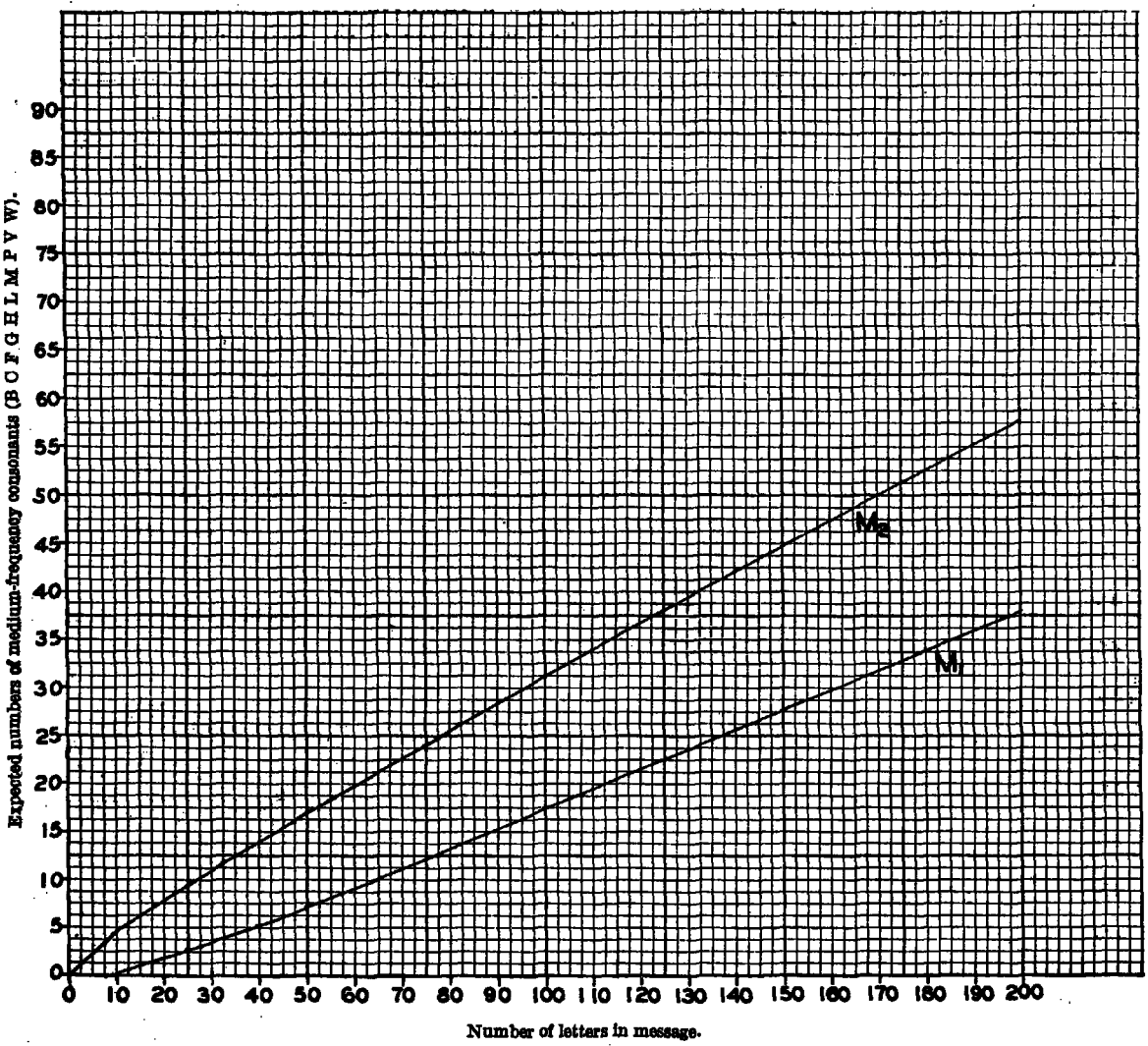


CHART NO. 3.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of medium-frequency consonants to be expected in messages of various lengths. (See Par. 13*d*.)

of letters, figures and other symbols, it is immediately apparent that the cryptogram is a substitution cipher.

i. Finally, it should be mentioned that there are certain kinds of cryptograms whose class cannot be determined by the method set forth in subparagraphs *b*, *c*, *d* above. These exceptions will be discussed in a subsequent section of this text.[1]

14. Determining whether a substitution cipher is monoalphabetic or polyalphabetic.—*a.* It will be remembered that a monoalphabetic substitution cipher is one in which a single cipher alphabet is employed throughout the whole message, that is, a given plain-text letter is invariably

[1] Par. 47.

represented throughout the message by one and the same letter in the cipher text. On the other hand, a polyalphabetic substitution cipher is one in which two or more cipher alphabets are employed within the same message; that is, a given plain-text letter may be represented by two or more different letters in the cipher text, according to some rule governing the selection of the equivalent to be used in each case. From this it follows that a single cipher letter may represent two or more different plain-text letters.

b. It is easy to see why and how the appearance of the uniliteral frequency distribution for a substitution cipher may be used to determine whether the cryptogram is monoalphabetic or polyalphabetic in character. The normal distribution presents marked crests and troughs by

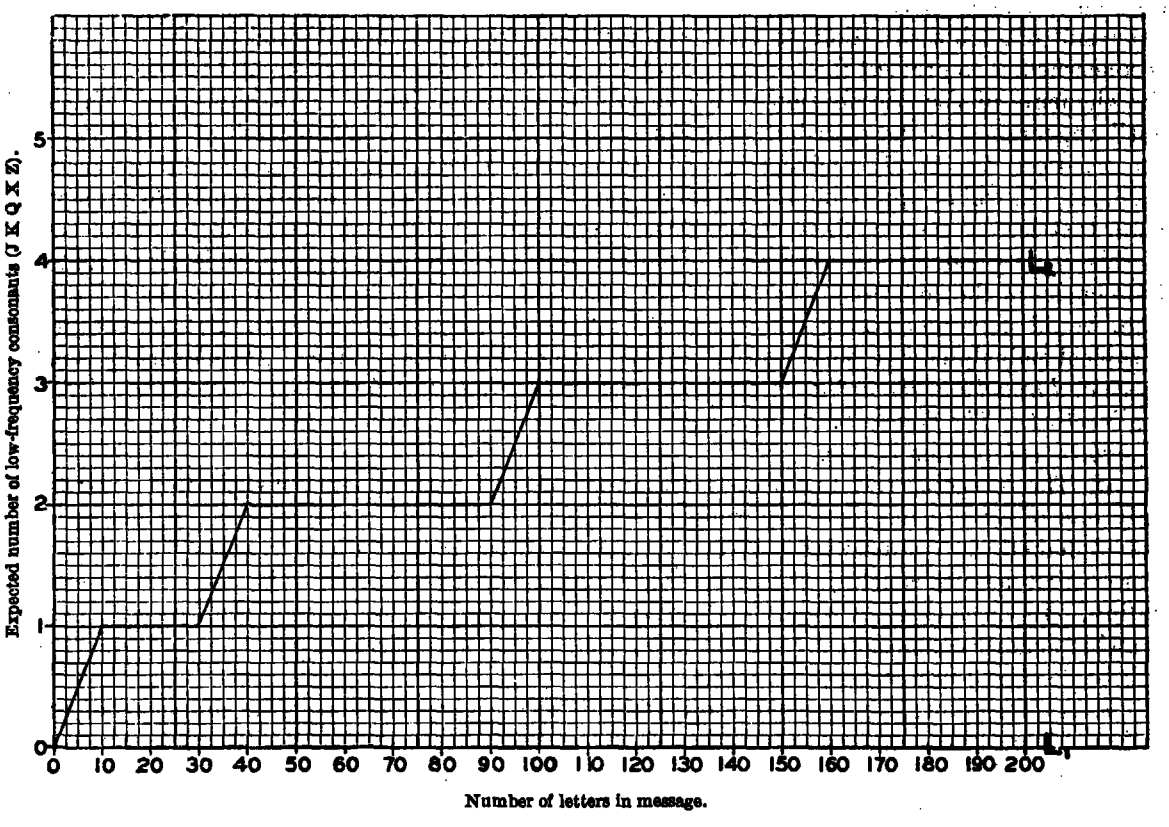


CHART No. 4.—Curves marking the lower and upper limits of the theoretical amount of deviation from the normal number of low-frequency consonants to be expected in messages of various lengths. (See Par. 13*d*.)

virtue of two circumstances. First, the elementary sounds which the symbols represent are used with greatly varying frequencies, it being one of the striking characteristics of every alphabetic language that its elementary sounds are used with greatly varying frequencies.[2] In the second place, except for orthographic aberrations peculiar to certain languages (conspicuously, English and French), each such sound is represented by the same symbol. It follows, therefore, that since in a monoalphabetic substitution cipher each different plain-text letter (=elementary sound) is represented by one and only one cipher letter (=elementary symbol), the uniliteral frequency distribution for such a cipher message must also exhibit the irregular crest and trough appearance of the normal distribution, but with only this important modification—*the absolute*

[2] The student who is interested in this phase of the subject may find the following reference of value: Zipf, G. K., *Selected Studies of the Principle of Relative Frequency in Language*, Cambridge, Mass., 1932.

positions of the crests and troughs will not be the same as in the normal. That is, the letters accompanying the crests and the troughs in the distribution for the cryptogram will be different from those accompanying the crests and the troughs in the normal distribution. But the marked irregularity of the distribution, the presence of accentuated crests and troughs, is in itself an indication that each symbol or cipher letter always represents the same plain-text letter in that cryptogram. Hence the general rule: *A marked crest and trough appearance in the uniliteral frequency distribution for a given cryptogram indicates that a single cipher alphabet is involved and constitutes one of the tests for a monoalphabetic substitution cipher.*

c. On the other hand, suppose that in a cryptogram each cipher letter represents several different plain-text letters. Some of them are of high frequency, others of low frequency. The net result of such a situation, so far as the uniliteral frequency distribution for the cryptogram is concerned, is to prevent the appearance of any marked crests and troughs and to tend to reduce the elements of the distribution to a more or less common level. This imparts a "flattened out" appearance to the distribution. For example, in a certain cryptogram of polyalphabetic construction, $K_c=E_p$, G_p, and J_p; $R_c=A_p$, D_p, and B_p; $X_c=O_p$, L_p, and F_p. The frequencies of K_c, R_c, and X_c will be approximately equal because the summations of the frequencies of the several plain-text letters each of these cipher letters represents at different times will be about equal. If this same phenomenon were true of all the letters of the cryptogram, it is clear that the frequencies of the 26 letters, when shown by means of the ordinary uniliteral frequency distribution, would show no striking differences and the distribution would have the flat appearance of a typical polyalphabetic substitution cipher. Hence, the general rule: *The absence of marked crests and troughs in the uniliteral frequency distribution indicates that two or more cipher alphabets are involved. The flattened-out appearance of the distribution constitutes one of the tests for a polyalphabetic substitution cipher.*

d. The foregoing test based upon the appearance of the frequency distribution constitutes only one of several means of determining whether a substitution cipher is monoalphabetic or polyalphabetic in composition. It can be employed in cases yielding frequency distributions from which definite conclusions can be drawn with more or less certainty by mere ocular examination. In those cases in which the frequency distributions contain insufficient data to permit drawing definite conclusions by such examination, certain statistical tests can be applied. These will be discussed in a subsequent text.

e. At this point, however, one additional test will be given because of its simplicity of application. It may be employed in testing messages up to 200 letters in length, it being assumed that in messages of greater length ocular examination of the frequency distribution offers little or no difficulty. This test concerns the *number of blanks* in the frequency distribution, that is, the number of letters of the alphabet which are entirely absent from the message. It has been found from statistical studies that rather definite "laws" govern the theoretically expected number of blanks in normal plain-text messages and in frequency distributions for cryptograms of different natures and of various sizes. The results of certain of these studies have been embodied in Chart 5.

f. This chart contains two curves. The one labeled *P* applies to the average number of blanks theoretically expected in frequency distributions based upon normal plain-text messages of the indicated lengths. The other curve, labeled *R*, applies to the average number of blanks theoretically expected in frequency distributions based upon perfectly *random* assortments of letters; that is, assortments such as would be found by random selection of letters out of a hat containing thousands of letters, all of the 26 letters of the alphabet being present in equal proportions, each letter being replaced after a record of its selection has been made. Such random assortments correspond to polyalphabetic cipher messages in which the number of cipher alpha-

bets is so large that if uniliteral frequency distributions are made of the letters, the distributions are practically identical with those which are obtained by random selections of letters out of a hat.

g. In using this chart, one finds the point of intersection of the vertical coordinate corresponding to the length of the message, with the horizontal coordinate corresponding to the observed number of blanks in the distribution for the message. If this point of intersection falls closer to curve *P* than it does to curve *R*, the number of blanks in the message approximates or corresponds more closely to the number theoretically expected in a plain-text message than it does to a random (cipher-text) message of the same length; therefore, this is evidence that the cryptogram is monoalphabetic. Conversely, if this point of intersection falls closer to curve *R*

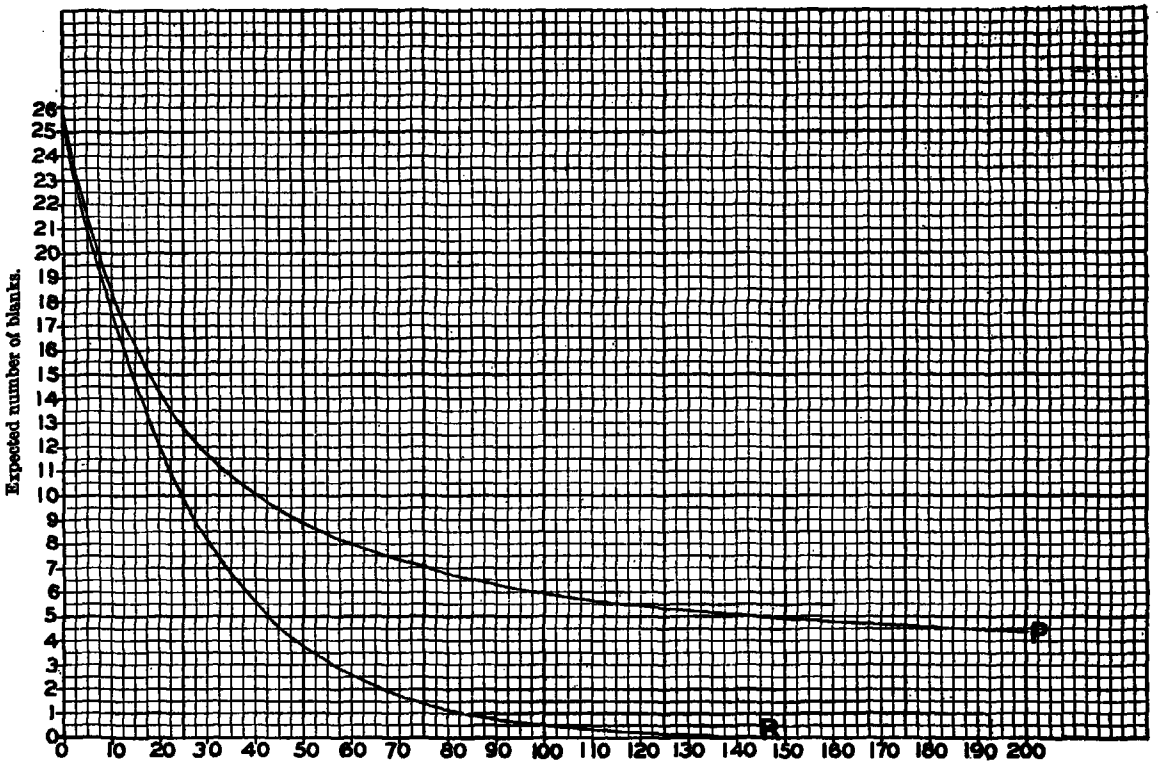


CHART NO. 5.—Curves showing the average number of blanks theoretically expected in distributions for plain text (P) and for random text (R) for messages of various lengths. (See Par. 14*f*.)

than to curve *P*, the number of blanks in the message approximates or corresponds more closely to the number theoretically expected in a random text than it does to a plain-text message of the same length; therefore, this is evidence that the cryptogram is polyalphabetic.

h. Practical examples of the use of this chart will be given in some of the illustrative messages to follow.

15. Determining whether the cipher alphabet is a standard, or a mixed cipher alphabet.—*a.* Assuming that the uniliteral frequency distribution for a given cryptogram has been made, and that it shows clearly that the cryptogram is a substitution cipher and is monoalphabetic in character, a consideration of the nature of standard cipher alphabets [3] almost makes it obvious how an inspection of the distribution will disclose whether the cipher alphabet involved is a standard cipher alphabet or a mixed cipher alphabet. If the crests and troughs of the distribu-

[3] See Sec. VIII, *Elementary Military Cryptography.*

tion occupy positions which correspond to the *relative* positions they occupy in the normal frequency distribution, then the cipher alphabet is a standard cipher alphabet. If this is not the case, then it is highly probable that the cryptogram has been prepared by the use of a mixed cipher alphabet.

b. A mechanical test may be applied in doubtful cases arising from lack of material available for study. Just what this test involves, and an illustration of its application will be given in the next section, using specific examples.

16. Determing whether the standard cipher alphabet is direct or reversed.—Assuming that the frequency distribution for a given cryptogram shows clearly that a standard cipher alphabet is involved, the determination as to whether the alphabet is direct or reversed can also be made by inspection, since the difference between the two is merely a matter of the *direction* in which the sequence of crests and troughs progresses—to the right, as in normal reading or writing, or the left. In a direct cipher alphabet the direction in which the crests and troughs of the distribution should be read is the normal direction, from left to right; in a reversed cipher alphabet this direction is reversed, from right to left.

SECTION V

UNILITERAL SUBSTITUTION WITH STANDARD CIPHER ALPHABETS

	Paragraph
Principles of solution by construction and analysis of the uniliteral frequency distribution	17
Theoretical example of solution	18
Practical example of solution by the frequency method	19
Solution by completing the plain-component sequence	20
Special remarks on the method of solution by completing the plain-component sequence	21
Value of mechanical solution as a short cut	22

17. Principles of solution by construction and analysis of the uniliteral frequency distribution.—*a.* Standard cipher alphabets are of two sorts, direct and reversed. The analysis of monoalphabetic cryptograms prepared by their use follows almost directly from a consideration of the nature of such alphabets. Since the cipher component of a standard cipher alphabet consists either of the normal sequence merely displaced 1, 2, 3, . . . intervals from the normal point of coincidence, or of the normal sequence proceeding in a reversed-normal direction, it is obvious that the uniliteral frequency distribution for a cryptogram prepared by means of such a cipher alphabet employed monoalphabetically will show crests and troughs whose *relative* positions and frequencies will be exactly the same as in the uniliteral frequency distribution for the plain text of that cryptogram. The only thing that has happened is that the whole set of crests and troughs of the distribution has been displaced to the right or left of the position it occupies in the distribution for the plain text; or else the successive elements of the whole set progress in the opposite direction. Hence, it follows that the correct determination of the plain-text value of the letter marking *any* crest or trough of the uniliteral frequency distribution will result at one stroke in the correct determination of the plain-text values of *all* the remaining 25 letters respectively marking the other crests and troughs in that distribution. Thus, having determined the value of a single element of the cipher component of the cipher alphabet, the values of all the remaining letters of the cipher component are automatically solved at one stroke. In more simple language, the correct determination of the value of a single letter of the cipher text automatically gives the values of the other 25 letters of the cipher text. The problem thus resolves itself into a matter of selecting that point of attack which will most quickly or most easily lead to the determination of the value of *one* cipher letter. The single word *identification* will hereafter be used for the phrase "determination of the value of a cipher letter"; to *identify* a cipher letter is to find its plain-text value.

b. It is obvious that the easiest point of attack is to assume that the letter marking the crest of greatest frequency in the frequency distribution for the cryptogram represents E_p. Proceeding from this initial point, the identifications of the remaining cipher letters marking the other crests and troughs are tentatively made on the basis that the letters of the cipher component proceed in accordance with the normal alphabetic sequence, either direct or reversed. If the actual frequency of each letter marking a crest or a trough approximates to a fairly close degree the normal theoretical frequency of the assumed plain-text equivalent, then the initial identification $\Theta_c = E_p$ may be *assumed to be correct* and therefore the derived identifications of the other cipher letters may be assumed to be correct. If the original starting point for assignment of plain-text values is not correct, or if the direction of "reading" the successive crests and troughs of the

distribution is not correct, then the frequencies of the other 25 cipher letters will not correspond to or even approximate the normal theoretical frequencies of their hypothetical plain-text equivalents on the basis of the initial identification. A new initial point, that is, a different cipher equivalent must then be selected to represent E_p; or else the direction of "reading" the crests and troughs must be reversed. This procedure, that is, the attempt to make the actual frequency relations exhibited by uniliteral frequency distribution for a given cryptogram conform to the theoretical frequency relations of the normal frequency distribution in an effort to solve the cryptogram, is referred to technically as "fitting the actual uniliteral frequency distribution for a cryptogram to the theoretical uniliteral frequency distribution for normal plain text", or, more briefly, as *"fitting the frequency distribution for the cryptogram to the normal frequency distribution"*, or, still more briefly, *"fitting the distribution to the normal."* In statistical work the expression commonly employed in connection with this process of fitting an actual distribution to a theoretical one is "testing the goodness of fit." The goodness of fit may be stated in various ways, mathematical in character.

c. In fitting the actual distribution to the normal, it is necessary to regard the cipher component (that is, the letters A . . . Z marking the successive crests and troughs of the distribution) as partaking of the nature of a wheel or sequence closing in upon itself, so that no matter with what crest or trough one starts, the spatial and frequency relations of the crests and troughs are constant. This manner of regarding the cipher component as being cyclic in nature is valid *because it is obvious that the relative positions and frequencies of the crests and troughs of any uniliteral-frequency distribution must remain the same regardless of what letter is employed as the initial point of the distribution.* Fig. 5 gives a clear picture of what is meant in this connection, as applied to the normal frequency distribution.

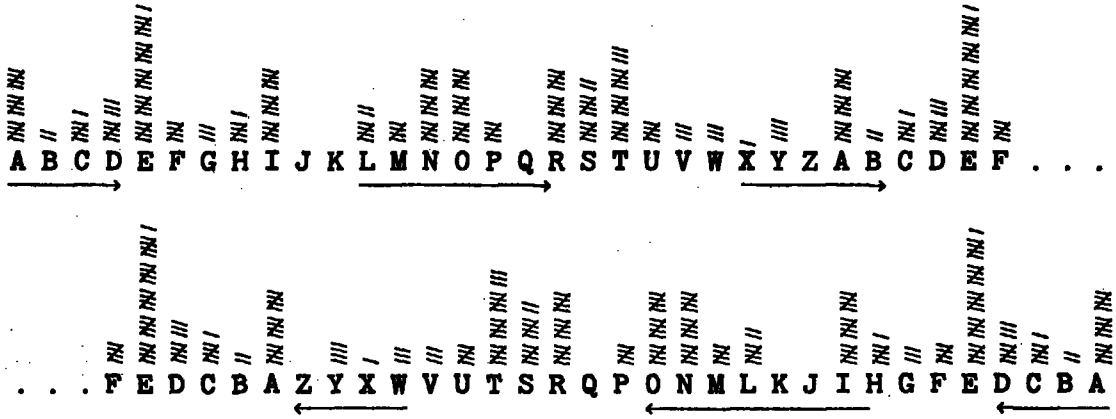


FIGURE 5.

d. In the third sentence of subparagraph *b*, the phrase "assumed to be correct" was advisedly employed in describing the results of the attempt to fit the distribution to the normal, because the final test of the goodness of fit in this connection (that is, of the correctness of the assignment of values to the crests and troughs of the distribution) is whether the *consistent* substitution of the plain-text values of the cipher characters in the cryptogram will yield intelligible plain text. If this is not the case, then no matter how close the approximation between actual and theoretical frequencies is, no matter how well the actual frequency distribution fits the normal, the only possible inferences are that (1) either the closeness of the fit is a pure coincidence in this case, and that another equally good fit may be obtained from the same data, or else (2) the cryptogram involves something more than simple monoalphabetic substitution by

means of a single standard cipher alphabet. For example, suppose a transposition has been applied in addition to the substitution. Then, although an excellent correspondence between the uniliteral frequency distribution and the normal frequency distribution has been obtained, the substitution of the cipher letters by their assumed equivalents will still not yield plain text. However, aside from such cases of double encipherment, instances in which the uniliteral frequency distribution may be easily fitted to the normal frequency distribution and in which at the same time an attempted simple substitution fails to yield intelligible text are rare. It may be said that, in practical operations whenever the uniliteral frequency distribution can be made to fit the normal frequency distribution, substitution of values will result in solution; and, as a corollary, whenever the uniliteral frequency distribution cannot be made to fit the normal frequency distribution, the cryptogram does not represent a case of simple, monoalphabetic substitution by means of a standard alphabet.

18. Theoretical example of solution.—*a.* The foregoing principles will become clearer by noting the cryptographing and solution of a theoretical example. The following message is to be cryptographed.

HOSTILE FORCE ESTIMATED AT ONE REGIMENT INFANTRY AND TWO PLATOONS CAVALRY MOVING SOUTH ON QUINNIMONT PIKE STOP HEAD OF COLUMN NEARING ROAD JUNCTION SEVEN THREE SEVEN COMMA EAST OF GREENACRE SCHOOL FIRED UPON BY OUR PATROLS STOP HAVE DESTROYED BRIDGE OVER INDIAN CREEK .

b. First, solely for purposes of demonstrating certain principles, the uniliteral frequency distribution for this message is presented in Figure 6.

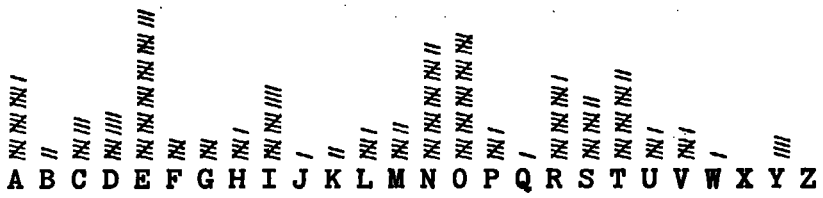


FIGURE 6.

c. Now let the foregoing message be cryptographed monoalphabetically by the following cipher alphabet, yielding the cryptogram and the frequency distribution shown below.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F

Plain	HOSTI	LEFOR	CEEST	IMATE	DATON	EREGI	MENTI	NFANT	RYAND
Cipher	NUYZO	RKLUX	IKKYZ	OSGZK	JGZUT	KXKMO	SKTZO	TLGTZ	XEGTJ
Plain	TWOPL	ATOON	SCAVA	LRYMO	VINGS	OUTHO	NQUIN	NIMON	TPIKE
Cipher	ZCUVR	GZUUT	YIGBG	RXESU	BOTMY	UAZNU	TWAOT	TOSUT	ZVOQK
Plain	STOPH	EADOF	COLUM	NNEAR	INGRO	ADJUN	CTION	SEVEN	THREE
Cipher	YZUVN	KGJUL	IURAS	TTKGX	OTMXU	GJPAT	IZOUT	YKBKT	ZNXKK
Plain	SEVEN	COMMA	EASTO	FGREE	NACRE	SCHOO	LFIRE	DUPON	BYOUR
Cipher	YKBKT	IUSSG	KGYZU	LMXKK	TGIXK	YINUU	RLOXK	JAVUT	HEUAX
Plain	PATRO	LSSTO	PHAVE	DESTR	OYEDB	RIDGE	OVERI	NDIAN	CREEK
Cipher	VGZXU	RYYZU	VNGBK	JKYZX	UEKJH	XOJMK	UBKXO	TJOGT	IXKKQ

CRYPTOGRAM

N U Y Z O R K L U X I K K Y Z O S G Z K J G Z U T K X K M O
S K T Z O T L G T Z X E G T J Z C U V R G Z U U T Y I G B G
R X E S U B O T M Y U A Z N U T W A O T T O S U T Z V O Q K
Y Z U V N K G J U L I U R A S T T K G X O T M X U G J P A T
I Z O U T Y K B K T Z N X K K Y K B K T I U S S G K G Y Z U
L M X K K T G I X K Y I N U U R L O X K J A V U T H E U A X
V G Z X U R Y Y Z U V N G B K J K Y Z X U E K J H X O J M K
U B K X O T J O G T I X K K Q

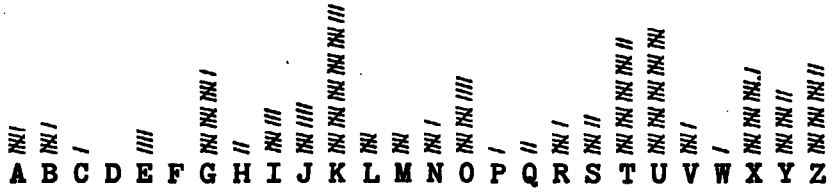


FIGURE 7

d. Let the student now compare Figs. 6 and 7, which have been superimposed in Fig. 8 for convenience in examination. Crests and troughs are present in both distributions; moreover their relative positions and frequencies have not been changed in the slightest particular. Only the absolute position of the sequence as a whole has been displaced six intervals to the right in Fig. 7, as compared with the absolute position of the sequence in Fig. 6.

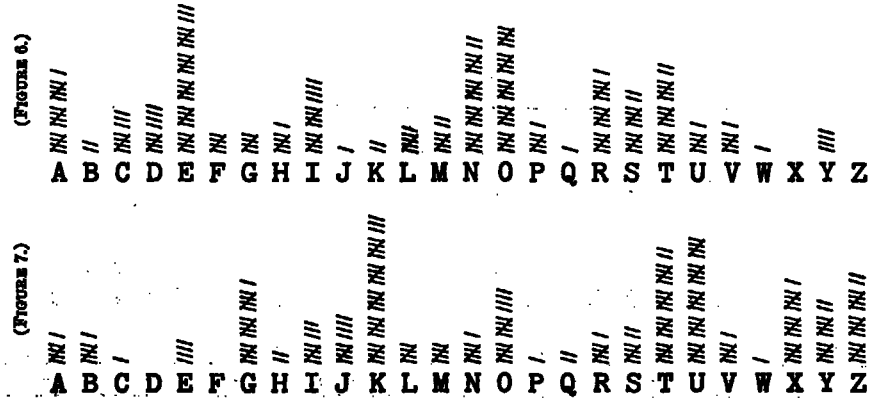


FIGURE 8.

e. If the two distributions are compared in detail the student will clearly understand how easy the solution of the cryptogram would be to one who knew nothing about how it was prepared. For example, the frequency of the highest crest, representing E_p in Fig. 6 is 28; at an interval of four letters before E_p there is another crest representing A_p with frequency 16. Between A and E there is a trough, representing the low-frequency letters B, C, D. On the other side of E, at an interval of four letters, comes another crest, representing I with frequency 14. Between E and I there is another trough, representing the low-frequency letters F, G, H. Compare these crests and troughs with their homologous crests and troughs in Fig. 7. In the latter, the letter K marks the highest crest in the distribution with a frequency of 28; four letters before K there is another crest, frequency 16, and four letters on the other side of K there is another crest, frequency

14. Troughs corresponding to B, C, D and F, G, H are seen at H, I, J and L, M, N in Fig. 7. In fact, the two distributions may be made to coincide exactly, by shifting the frequency distribution for the cryptogram six intervals to the left with respcet to the distribution for the equivalent plain-text message, as shown herewith.

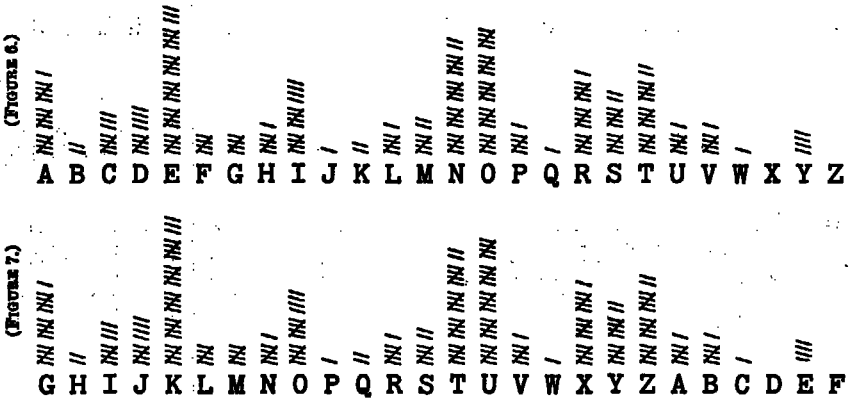


FIGURE 9.

f. Let us suppose now that nothing is known about the cryptographing process, and that only the cryptogram and its uniliteral frequency distribution is at hand. It is clear that simply bearing in mind the spatial relations of the crests and troughs in a normal frequency distribution would enable the cryptanalyst to fit the distribution to the normal in this case. He would naturally first assume that $G_c=A_p$, from which it would follow that if a direct standard alphabet is involved, $H_c=B_p$, $I_c=C_p$, and so on, yielding the following (tentative) deciphering alphabet:

Cipher	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Plain	U V W X Y Z A B C D E F G H I J K L M N O P Q R S T

g. Now comes the final test: If these assumed values are substituted in the cipher text, the plain text immediately appears. Thus:

N	U	Y	Z	O	R	K	L	U	X	etc.
H	O	S	T	I	L	E	F	O	R	etc.
I	K	K	Y	Z	O	S	G	Z	K	etc.
C	E	E	S	T	I	M	A	T	E	etc.
J	G	Z	U	T	D	A	T	O	N	etc.

h. It should be clear, therefore, that the selection of G_c to represent A_p in the cryptographing process has absolutely no effect upon the relative spatial and frequency relations of the crests and troughs of the frequency distribution for the cryptogram. If Q_c had been selected to represent A_p, these relations would still remain the same, the whole series of crests and troughs being merely displaced further to the right of the positions they occupy when $G_c=A_p$.

19. Practical example of solution by the frequency method.—*a. The case of direct standard alphabet ciphers.*—(1) The following cryptogram is to be solved by applying the foregoing principles:

I	B	M	Q	O	P	B	I	U	O	M	B	B	G	A	J	C	Z	O	F	M	U	U	Q	B	A	J	C	Z	O
Z	W	I	L	N	Q	T	T	M	L	E	Q	B	P	U	I	Z	K	P	Q	V	O	Q	V	N	I	V	B	Z	G

(2) From the presence of repetitions and so many low-frequency letters such as B, Q, and Z it is at once suspected that this is a substitution cipher. But to illustrate the steps that must be taken in difficult cases in order to be certain in this respect, a uniliteral frequency distribution

is constructed, and then reference is made to charts 1 to 4 to note whether the actual numbers of vowels, high, medium, and low-frequency consonants fall inside or outside the areas delimited by the respective curves.

FIGURE 10a.

Letters	Frequency	Position with respect to areas delimited by curves
Vowels (A E I O U Y)	17	Outside, chart 1.
High-frequency Consonants (D N R S T)	4	Outside, chart 2.
Medium-frequency Consonants (B C F G H L M P V W)	25	Outside, chart 3.
Low-frequency Consonants (J W Q X Z)	14	Outside, chart 4.
Total	60	

(3) All four points falling quite outside the areas delimited by the curves applicable to these four classes of letters, the cryptogram is clearly a substitution cipher.

(4) The appearance of the frequency distribution, with marked crests and troughs, indicates that the cryptogram is probably monoalphabetic. Reference is now made to Chart 5. The message has 60 letters and 6 blanks. The point of intersection on the chart is closer to curve *P* than it is to curve *R*; therefore, this is additional evidence that the message is probably monoalphabetic.

(5) The next step is to determine whether a standard or a mixed cipher alphabet is involved. This is done by studying the positions and the sequence of crests and troughs in the frequency distribution, and trying to fit the distribution to the normal.

(6) The first assumption to be made is that a direct standard is involved. The highest crest in the distribution is marked by B_c. Let it be assumed that $B_c = E_p$. Then C_c, D_c, E_c, . . . $= F_p$, G_p, H_p, . . ., respectively; thus:

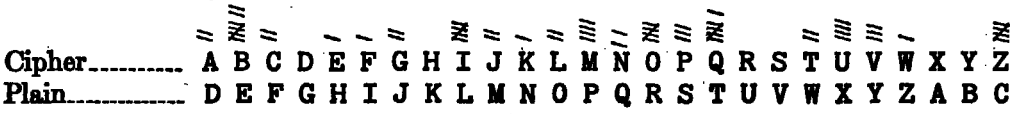


FIGURE 10b.

At first glance the approximation to the expected frequencies seems fair, especially in the region F G H I J K_p and R S T_p. But there are too many occurrences of L_p, P_p, X_p and C_p and too few occurrences of A_p, I_p, N_p, O_p. Moreover, if a substitution is attempted on this basis, the following is obtained for the first two cipher groups:

Cipher	I	B	M	Q	O		P	B	I	U	O
"Plain text"	L	E	P	T	R		S	E	L	X	R

This is certainly not plain text and it seems clear that B_c is not E_p. A different assumption will have to be made.

(7) Suppose $Q_c = E_p$. Going through the same steps as before, again no satisfactory results are obtained. Further trials[1] are made along the same lines, until the assumption $M_c = E_p$ is tested.

[1] It is unnecessary, of course, to write out the alphabets as shown in Figs. 10*b* and *c* when testing assumptions. This is usually all done mentally.

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R

FIGURE 10c.

(8) The fit in this case is quite good; possibly there are too many occurrences of G_p and M_p and two few of E_p, O_p and S_p. But the final test remains: trial of the substitution alphabet on the cryptogram itself. This is immediately done and the results are as follows:

Cryptogram	I	B	M	Q	O	P	B	I	U	O	N	B	B	G	A	J	C	Z	O	F
Plain text	A	T	E	I	G	H	T	A	M	G	E	T	T	Y	S	B	U	R	G	X
Cryptogram	M	U	U	Q	B	A	J	C	Z	O	Z	W	I	L	N	Q	T	T	M	L
Plain text	E	M	M	I	T	S	B	U	R	G	R	O	A	D	F	I	L	L	E	D
Cryptogram	E	Q	B	P	U	I	Z	K	P	Q	V	O	Q	V	N	I	V	B	Z	G
Plain text	W	I	T	H	M	A	R	C	H	I	N	G	I	N	F	A	N	T	R	Y

AT EIGHT AM GETTYSBURG-EMMITSBURG ROAD FILLED WITH MARCHING INFANTRY.

(9) It is always advisable to note the specific key. In this case the correspondence between any plain-text letter and its cipher equivalent will indicate the key. Although other conventions are possible, and equally valid, it is usual, however, to indicate the key by noting the cipher equivalent of A_p. In this case $A_p = I_c$.

b. The case of reversed standard alphabet ciphers.—

(1) Let the following cryptogram and its frequency distribution be studied.

I	P	E	A	C	B	P	I	W	C	E	P	P	K	Q	H	O	R	C	L	
E	W	W	A	P	Q	H	O	R	C	R	U	I	F	D	A	X	X	E	F	
M	A	P	B	W	I	R	G	B	A	V	C	A	V	D	I	V	P	R	K	

(2) The preliminary steps illustrated above, under subpar. *a* (1) to (4) inclusive, in connection with the test for class and monoalphabeticity, will here be omitted, since they are exactly the same in nature. The result is that the cryptogram is obviously a substitution cipher and is monoalphabetic.

(3) Assuming that it is not known whether a direct or a reversed standard alphabet is involved, attempts are at once made to fit the frequency distribution to the normal direct sequence. If the student will try them he will soon find out that these are unsuccessful. All this takes but a few minutes.

(4) The next logical assumption is now made, viz, that the cipher alphabet is a reversed standard alphabet. When on this basis E_c is assumed to be E_p, the distribution can readily be fitted to the normal, practically every crest and trough in the actual distribution corresponding to a crest or trough in the expected distribution.

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	I	H	G	F	E	D	C	B	A	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L	K	J

FIGURE 10d.

(5) When the substitution is made in the cryptogram, the following is obtained.

Cryptogram	I	P	E	A	C	B	P	I	W	C	E	P	P	K	Q	.	.	.		
Plain text	A	T	E	I	G	H	T	A	M	G	E	T	T	Y	S	.	.	.		

(6) The plain-text message is identical with that under paragraph *a*. The specific key in this case is also $A_p = I_c$. If the student will compare the frequency distributions in the two cases,

he will note that the relative positions and extensions of the crests and troughs are identical; they merely progress in opposite directions.

20. Solution by completing the plain-component sequence.—*a. The case of direct standard alphabet ciphers.*—(1) The foregoing method of analysis, involving as it does the construction of a uniliteral frequency distribution, was termed a *solution by the frequency method* because it involves the construction of a frequency distribution and its study. There is, however, another method which is much more rapid, almost wholly mechanical, and which, moreover, does not necessitate the construction or study of any frequency distribution whatever. An understanding of the method follows from a consideration of the method of encipherment of a message by the use of a single, direct standard cipher alphabet.

(2) Note the following encipherment:

Message..... REPEL INVADING CAVALRY

ENCIPHERING ALPHABET

Plain.....	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher.....	G H I J K L M N O P Q R S T U V W X Y Z A B C D E F

ENCIPHERMENT

Plain text.....	R E P E L I N V A D I N G C A V A L R Y
Cryptogram....	X K V K R O T B G J O T M I G B G R X E

CRYPTOGRAM

X K V K R O T B G J O T M I G B G R X E

(3) The enciphering alphabet shown above represents a case wherein the sequence of letters of both components of the cipher alphabet is the normal sequence, with the sequence forming the cipher component merely shifted six intervals in retard (or 20 intervals in advance) of the position it occupies in the normal alphabet. If, therefore, two strips of paper bearing the letters of the normal sequence, equally spaced, are regarded as the two components of the cipher alphabet and are juxtaposed at all of the 25 possible points of coincidence, it is obvious that one of these 25 juxtapositions *must* correspond to the actual juxtaposition shown in the enciphering alphabet directly above.[2] It is equally obvious that if a record were kept of the results obtained by applying the values given at each juxtaposition to the letters of the cryptogram, one of these results would yield the plain text of the cryptogram.

(4) Let the work be systematized and the results set down in an orderly manner for examination. It is obviously unnecessary to juxtapose the two components so that $A_c=A_p$, for on the assumption of a direct standard alphabet, juxtaposing two direct normal components at their normal point of coincidence merely yields plain text. The next possible juxtaposition, therefore, is $A_c=B_p$. Let the juxtaposition of the two sliding strips therefore be $A_c=B_p$, as shown here:

Plain.....	ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher.....	ABCDEFGHIJKLMNOPQRSTUVWXYZ

The values given by this juxtaposition are substituted for the first 20 letters of the cryptogram and the following results are obtained.

Cryptogram.....	X K V K R O T B G J O T M I G B G R X E
1st Test—"Plain text"...	Y L W L S P U C H K P U N J H C H S Y F

[2] One of the strips should bear the sequence repeated. This permits juxtaposing the two sequences at all 26 possible points of coincidence so as to have a complete cipher alphabet showing at all times.

This certainly is not intelligible text; obviously, the two components were not in the position indicated in this first test. The cipher component is therefore slid one interval to the right, making $A_c = C_p$, and a second test is made. Thus

Plain	ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cryptogram	X K V K R O T B G J O T M I G B G R X E
2d Test—"Plain text"	Z M X M T Q V D I L Q V O K I D I T Z G

Neither does the second test result in disclosing any plain text. But, if the results of the two tests are studied a phenomenon that at first seems quite puzzling comes to light. Thus, suppose the results of the two tests are superimposed in this fashion.

Cryptogram	X K V K R O T B G J O T M I G B G R X E
1st Test—"Plain text"	Y L W L S P U C H K P U N J H C H S Y F
2nd Test—"Plain text"	Z M X M T Q V D I L Q V O K I D I T Z G

(5) Note what has happened. The net result of the two experiments was merely to continue the normal sequence begun by the cipher letters at the heads of the several *columns*. It is obvious that if the normal sequence is completed in each column *the results will be exactly the same as though the whole set of 25 possible tests had actually been performed.* Let the columns therefore be completed, as shown in Fig. 11.

X	K	V	K	R	O	T	B	G	J	O	T	M	I	G	B	G	R	X	E
Y	L	W	L	S	P	U	C	H	K	P	U	N	J	H	C	H	S	Y	F
Z	M	X	M	T	Q	V	D	I	L	Q	V	O	K	I	D	I	T	Z	G
A	N	Y	N	U	R	W	E	J	M	R	W	P	L	J	E	J	U	A	H
B	O	Z	O	V	S	X	F	K	N	S	X	Q	M	K	F	K	V	B	I
C	P	A	P	W	T	Y	G	L	O	T	Y	R	N	L	G	L	W	C	J
D	Q	B	Q	X	U	Z	H	M	P	U	Z	S	O	M	H	M	X	D	K
E	R	C	R	Y	V	A	I	N	Q	V	A	T	P	N	I	N	Y	E	L
F	S	D	S	Z	W	B	J	O	R	W	B	U	Q	O	J	O	Z	F	M
G	T	E	T	A	X	C	K	P	S	X	C	V	R	P	K	P	A	G	N
H	U	F	U	B	Y	D	L	Q	T	Y	D	W	S	Q	L	Q	B	H	O
I	V	G	V	C	Z	E	M	R	U	Z	E	X	T	R	M	R	C	I	P
J	W	H	W	D	A	F	N	S	V	A	F	Y	U	S	N	S	D	J	Q
K	X	I	X	E	B	G	O	T	W	B	G	Z	V	T	O	T	E	K	R
L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T
N	A	L	A	H	E	J	R	W	Z	E	J	C	Y	W	R	W	H	N	U
O	B	M	B	I	F	K	S	X	A	F	K	D	Z	X	S	X	I	O	V
P	C	N	C	J	G	L	T	Y	B	G	L	E	A	Y	T	Y	J	P	W
Q	D	O	D	K	H	M	U	Z	C	H	M	F	B	Z	U	Z	K	Q	X
*R	E	P	E	L	I	N	V	A	D	I	N	G	C	A	V	A	L	R	Y
S	F	Q	F	M	J	O	W	B	E	J	O	H	D	B	W	B	M	S	Z
T	G	R	G	N	K	P	X	C	F	K	P	I	E	C	X	C	N	T	A
U	H	S	H	O	L	Q	Y	D	G	L	Q	J	F	D	Y	D	O	U	B
V	I	T	I	P	M	R	Z	E	H	M	R	K	G	E	Z	E	P	V	C
W	J	U	J	Q	N	S	A	F	I	N	S	L	H	F	A	F	Q	W	D

FIGURE 11.

An examination of the successive horizontal lines of the diagram discloses *one and only one* line of plain text, that marked by the asterisk and reading R E P E L I N V A D I N G C A V A L R Y.

(6) Since each column in Fig. 11 is nothing but a normal sequence, it is obvious that instead of laboriously writing down these columns of letters every time a cryptogram is to be examined, it would be more convenient to prepare a set of strips each bearing the normal sequence doubled (to permit complete coincidence for an entire alphabet at any setting), and have them available for examining any future cryptograms. In using such a set of sliding strips in order to solve a cryptogram prepared by means of a single direct standard cipher alphabet, or to make a test to determine whether a cryptogram has been so prepared, it is only necessary to "set up" the letters of the cryptogram on the strips, that is, align them in a single row across the strips (by sliding the individual strips up or down). The successive horizontal lines, called *generatrices* (singular, *generatrix*), are then examined in a search for intelligible text. If the cryptogram really belongs to this simple type of cipher, one of the generatrices will exhibit intelligible text all the way across; this text will practically invariably be the plain text of the message. This method of analysis may be termed *a solution by completing the plain-component sequence.* Sometimes it is referred to as "running down" the sequence. The principle upon which the method is based constitutes one of the cryptanalyst's most valuable tools.[8]

b. The case of reversed standard alphabets.—(1) The method described under subpar. *a* may also be applied, in slightly modified form, in the case of a cryptogram enciphered by a single reversed standard alphabet. The basic principles are identical in the two cases.

(2) To show this it is necessary to experiment with two sliding components as before, except that in this case one of the components must be a reversed normal sequence, the other, a direct normal sequence.

(3) Let the two components be juxtaposed A to A, as shown below, and then let the resultant values be substituted for the letters of the cryptogram. Thus:

CRYPTOGRAM

P C R C V Y T L G D Y T A E G L G V P I

Plain..... ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher..... ZYXWVUTSRQPONMLKJIHGFEDCBA

Cryptogram.....	P	C	R	C	V	Y	T	L	G	D	Y	T	A	E	G	L	G	V	P	I
1st Test—"Plain text"...	L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S

(4) This does not yield intelligible text, and therefore the reversed component is slid one space forward and a second test is made. Thus:

Plain..... ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher..... ZYXWVUTSRQPONMLKJIHGFEDCBA

Cryptogram.....	P	C	R	C	V	Y	T	L	G	D	Y	T	A	E	G	L	G	V	P	I
2d Test—"Plain text"...	M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T

(5) Neither does the second test yield intelligible text. But let the results of the two tests be superimposed. Thus:

Cryptogram.....	P	C	R	C	V	Y	T	L	G	D	Y	T	A	E	G	L	G	V	P	I
1st Test—"Plain text"...	L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
2d Test—"Plain text"...	M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T

[8] It is recommended that the student prepare a set of 25 strips ¼ by ½ by 15 inches, made of well-seasoned wood, and glue alphabet strips to the wood. The alphabet on each strip should be a double or repeated alphabet with all letters equally spaced.

(6) It is seen that the letters of the "plain text" given by the *second* trial are merely the continuants of the normal sequences initiated by the letters of the "plain text" given by the first trial. If these sequences are "run down"—that is, completed within the columns—the results must obviously be the same as though successive tests exactly similar to the first two were applied to the cryptogram, using one reversed normal and one direct normal component. If the cryptogram has really been prepared by means of a single reversed standard alphabet, one of the generatrices of the diagram that results from completing the sequences *must* yield intelligible text.

(7) Let the diagram be made, or better yet, if the student has already at hand the set of sliding strips referred to in the footnote to page 36, let him "set up" the letters given by the *first* trial. Fig. 12 shows the diagram and indicates the plain-text generatrix.

P	C	R	C	V	Y	T	L	G	D	Y	T	A	E	G	L	G	V	P	I
L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T
N	A	L	A	H	E	J	R	W	Z	E	J	C	Y	W	R	W	H	N	U
O	B	M	B	I	F	K	S	X	A	F	K	D	Z	X	S	X	I	O	V
P	C	N	C	J	G	L	T	Y	B	G	L	E	A	Y	T	Y	J	P	W
Q	D	O	D	K	H	M	U	Z	C	H	M	F	B	Z	U	Z	K	Q	X
*R	E	P	E	L	I	N	V	A	D	I	N	G	C	A	V	A	L	R	Y
S	F	Q	F	M	J	O	W	B	E	J	O	H	D	B	W	B	M	S	Z
T	G	R	G	N	K	P	X	C	F	K	P	I	E	C	X	C	N	T	A
U	H	S	H	O	L	Q	Y	D	G	L	Q	J	F	D	Y	D	O	U	B
V	I	T	I	P	M	R	Z	E	H	M	R	K	G	E	Z	E	P	V	C
W	J	U	J	Q	N	S	A	F	I	N	S	L	H	F	A	F	Q	W	D
X	K	V	K	R	O	T	B	G	J	O	T	M	I	G	B	G	R	X	E
Y	L	W	L	S	P	U	C	H	K	P	U	N	J	H	C	H	S	Y	F
Z	M	X	M	T	Q	V	D	I	L	Q	V	O	K	I	D	I	T	Z	G
A	N	Y	N	U	R	W	E	J	M	R	W	P	L	J	E	J	U	A	H
B	O	Z	O	V	S	X	F	K	N	S	X	Q	M	K	F	K	V	B	I
C	P	A	P	W	T	Y	G	L	O	T	Y	R	N	L	G	L	W	C	J
D	Q	B	Q	X	U	Z	H	M	P	U	Z	S	O	M	H	M	X	D	K
E	R	C	R	Y	V	A	I	N	Q	V	A	T	P	N	I	N	Y	E	L
F	S	D	S	Z	W	B	J	O	R	W	B	U	Q	O	J	O	Z	F	M
G	T	E	T	A	X	C	K	P	S	X	C	V	R	P	K	P	A	G	N
H	U	F	U	B	Y	D	L	Q	T	Y	D	W	S	Q	L	Q	B	H	O
I	V	G	V	C	Z	E	M	R	U	Z	E	X	T	R	M	R	C	I	P
J	W	H	W	D	A	F	N	S	V	A	F	Y	U	S	N	S	D	J	Q
K	X	I	X	E	B	G	O	T	W	B	G	Z	V	T	O	T	E	K	R

FIGURE 12.

(8) The only difference in procedure between this case and the preceding one (where the cipher alphabet was a direct standard alphabet) is that the letters of the cipher text are first "deciphered" by means of *any* reversed standard alphabet and then the columns are "run down", according to the normal A B C . . . Z sequence. For reasons which will become apparent very soon, the first step in this method is technically termed *converting the cipher letters into their plain-component equivalents;* the second step is the same as before, *viz, completing the plain-component sequence.*

21. Special remarks on the method of solution by completing the plain-component sequence.— *a.* The terms employed to designate the steps in the solution set forth in Par. 20*b*, viz, "converting the cipher letters into their plain-component equivalents" and "completing the plain-component sequence", accurately describe the process. Their meaning will become more clear as the student progresses with the work. It may be said that whenever the plain component of a cipher alphabet is a *known* sequence, no matter how it is composed, the difficulty and time required to solve any cryptogram involving the use of that plain component is practically cut in half. *In some cases this knowledge facilitates, and in other cases is the only thing that makes possible the solution of a very short cryptogram that might otherwise defy solution.* Later on an example will be given to illustrate what is meant in this regard.

b. The student should take note, however, of two qualifying expressions that were employed in a preceding paragraph to describe the results of the application of the method. It was stated that "one of the generatrices will exhibit intelligible text *all the way across;* this text will *practically invariably* be the plain text." Will there ever be a case in which more than one generatrix will yield intelligible text throughout its extent? That obviously depends almost entirely on the number of letters that are aligned to form a generatrix. If a generatrix contains but a very few letters, only five, for example, it may happen as a result of pure chance that there will be two or more generatrices showing what might be "intelligible text." Note in Fig. 11, for example, that there are several cases in which 3-letter and 4-letter English words (ANY, VAIN, GOT, TIP, etc.) appear on generatrices that are not correct, these words being formed by pure chance. But there is not a single case, in this diagram, of a 5-letter or longer word appearing fortuitously, because obviously the longer the word the smaller the probability of its appearance purely by chance; and the probability that two generatrices of 15 letters each will both yield intelligible text along their entire length is exceedingly remote, so remote, in fact, that in practical cryptography such a case may be considered nonexistent.[4]

c. The student should observe that in reality there is no difference whatsoever in principle between the two methods presented in subpars. *a* and *b* of Par. 20. In the former the preliminary step of converting the cipher letters into their plain-component equivalents is apparently not present but in reality it is there. The reason for its apparent absence is that in that case the plain component of the cipher alphabet is identical in all respects with the cipher component, so that the cipher letters require no conversion, or, rather, they are identical with the equivalents that would result if they were converted on the basis $A_c = A_p$. In fact, if the solution process had been arbitrarily initiated by converting the cipher letters into their plain-component equivalents at the setting $A_c = O_p$, for example, and the cipher component slid one interval to the right thereafter, the results of the first and second tests of Par. 20*a* would be as follows:

Cryptogram	X	K	V	K	R	O	T	B	G	J	O	T	M	I	G	B	G	R	X	E
1st Test—"Plain text"	L	Y	J	Y	F	C	H	P	U	X	C	H	A	W	U	P	U	F	L	S
2nd Test—"Plain text"	M	Z	K	Z	G	D	I	Q	V	Y	D	I	B	X	V	Q	V	G	M	T

Thus, the foregoing diagram duplicates in every particular the diagram resulting from the first two tests under Par. 20*b*: a first line of cipher letters, a second line of letters derived from them but showing externally no relationship with the first line, and a third line derived immediately from the second line by continuing the direct normal sequence. This point is brought to attention only for the purpose of showing that a single, broad principle is the basis of the general method of solution by completing the plain-component sequence, and once the student has this firmly in

[4] A person with patience and an inclination toward the curiosities of the science might construct a text of 15 or more letters which would yield two "intelligible" texts on the plain-component completion diagram.

mind he will have no difficulty whatsoever in realizing when the principle is applicable, what a powerful cryptanalytic tool it can be, and what results he may expect from its application in specific instances.

d. In the two foregoing examples of the application of the principle, the plain component was a normal sequence but it should be clear to the student, if he has grasped what has been said in the preceding subparagraph, that this component may be a mixed sequence which, if known (that is, if the sequence of letters comprising the sequence is known to the cryptanalyst), can be handled just as readily as can a plain component that is a normal sequence.

e. It is entirely immaterial at what points the plain and the cipher components are juxtaposed in the preliminary step of converting the cipher letters into their plain-component equivalents. For example, in the case of the reversed alphabet cipher solved in Par. 20*b*, the two components were arbitrarily juxtaposed to give the value A=A, but they might have been juxtaposed at any of the other 25 possible points of coincidence without in any way affecting the final result, *viz*, the production of one plain-text generatrix in the completion diagram.

22. Value of mechanical solution as a short cut.—*a.* It is obvious that the very first step the student should take in his attempts to solve an unknown cryptogram that is obviously a substitution cipher is to try the mechanical method of solution by completing the plain-component sequence, using the normal alphabet, first direct, then reversed. This takes only a very few minutes and is conclusive in its results. It saves the labor and trouble of constructing a frequency distribution in case the cipher is of this simple type. Later on it will be seen how certain variations of this simple type may also be solved by the application of this method. Thus, a very easy short cut to solution is afforded, which even the experienced cryptanalyst never overlooks in his first attack on an unknown cipher.

b. It is important now to note that *if neither of the two foregoing attempts is successful in bringing plain text to light and the cryptogram is quite obviously monoalphabetic in character, the cryptanalyst is warranted in assuming that the cryptogram involves a mixed cipher alphabet.*[5] The steps to be taken in attacking a cipher of the latter type will be discussed in the next section.

[5] There is but one other possibility, already referred to under Par. 17*d*, which involves the case where transposition and monoalphabetic substitution processes have been applied in successive steps. This is unusual, however, and will be discussed in its proper place.

SECTION VI

UNILITERAL SUBSTITUTION WITH MIXED CIPHER ALPHABETS

	Paragraph
Basic reason for the low degree of cryptographic security afforded by monoalphabetic cryptograms involving standard cipher alphabets	23
Preliminary steps in the analysis of a monoalphabetic, mixed-alphabet cryptogram	24
Further data concerning normal plain text	25
Preparation of the work sheet	26
Triliteral-frequency distributions	27
Classifying the cipher letters into vowels and consonants	28
Further analysis of the letters representing vowels and consonants	29
Substituting deduced values in the cryptogram	30
Completing the solution	31
General remarks on the foregoing solution	32
The "probable-word" method; its value and applicability	33
Solution of additional cryptograms produced by the same cipher component	34

23. Basic reason for the low degree of cryptographic security afforded by monoalphabetic cryptograms involving standard cipher alphabets.—The student has seen that the solution of monoalphabetic cryptograms involving standard cipher alphabets is a very easy matter. Two methods of analysis were described, one involving the construction of a frequency distribution, the other not requiring this kind of tabulation, being almost mechanical in nature and correspondingly rapid. In the first of these two methods it was necessary to make a correct assumption as to the value of but one of the 26 letters of the cipher alphabet and the values of the remaining 25 letters at once become known; in the second method it was not necessary to assume a value for even a single cipher letter. The student should understand what constitutes the basis of this situation, *viz*, the fact that the two components of the cipher alphabet are composed of *known sequences*. What if one or both of these components are, for the cryptanalyst, *unknown sequences*? In other words, what difficulties will confront the cryptanalyst if the cipher component of the cipher alphabet is a mixed sequence? Will such an alphabet be solvable as a whole at one stroke, or will it be necessary to solve its values individually? Since the determination of the value of one cipher letter in this case gives no direct clues to the value of any other letter, it would seem that the solution of such a cipher should involve considerably more analysis and experiment than has the solution of either of the two types of ciphers so far examined occasioned. A typical example will be studied.

24. Preliminary steps in the analysis of a monoalphabetic, mixed alphabet cryptogram.—*a*. Note the following cryptogram:

SFDZF IOGHL PZFGZ DYSPF HBZDS GVHTF UPLVD FGYVJ VFVHT GADZZ AITYD
ZYFZJ ZTGPT VTZBD VFHTZ DFXSB GIDZY VTXOI YVTEF VMGZZ THLLV XZDFM
HTZAI TYDZY BDVFH TZDFK ZDZZJ SXISG ZYGAV FSLGZ DTHHT CDZRS VTYZD
OZFFH TZAIT YDZYG AVDGZ ZTKHI TYZYS DZGHU ZFZTG UPGDI XWGHX ASRUZ
DFUID EGHTV EAGXX

b. A casual inspection of the text discloses the presence of several long repetitions as well as of many letters of normally low frequency, such as F, G, V, X, and Z; on the other hand, letters of

normally high frequency, such as the vowels, and the consonants N and R, are relatively scarce. The cryptogram is obviously a substitution cipher and the usual mechanical tests for determining whether it is possibly of the monoalphabetic, standard-alphabet type are applied. The results being negative, a uniliteral frequency distribution is immediately constructed and is as shown in Figure 13.

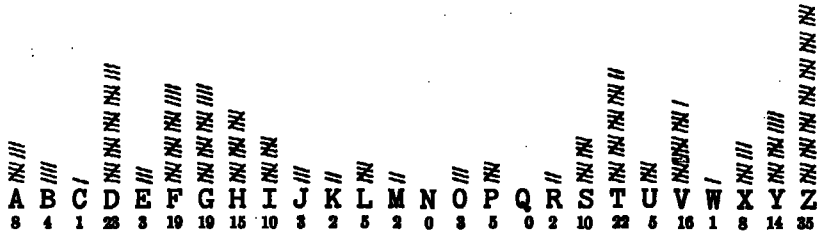


FIGURE 13

c. The fact that the frequency distribution shows very marked crests and troughs means that the cryptogram is undoubtedly monoalphabetic; the fact that it has already been tested (by the method of completing the plain-component sequence) and found not to be of the monoalphabetic, standard-alphabet type, indicates with a high degree of probability that it involves a mixed cipher alphabet. A few moments might be devoted to making a careful inspection of the distribution to insure that it cannot be made to fit the normal; the object of this would be to rule out the possibility that the text resulting from substitution by a standard cipher alphabet had not subsequently been transposed. But this inspection in this case is hardly necessary, in view of the presence of long repetitions in the message.[1] (See Par. 13*g*.)

d. One might, of course, attempt to solve the cryptogram by applying the simple principles of frequency. One might, in other words, assume that Z_c (the letter of greatest frequency) represents E_p, D_c (the letter of next greatest frequency) represents T_p, and so on. If the message were long enough this simple procedure might more or less quickly give the solution. But the message is relatively short and many difficulties would be encountered. Much time and effort would be expended unnecessarily, because it is hardly to be expected that in a message of only 235 letters the relative order of frequency of the various cipher letters should exactly coincide with, or even closely approximate the relative order of frequency of letters of normal plain text found in a count of 50,000 letters. *It is to be emphasized that the beginner must repress the natural tendency to place too much confidence in the generalized principles of frequency and to rely too much upon them.* It is far better to bring into effective use certain other data concerning normal plain text which thus far have not been brought to notice.

25. Further data concerning normal plain text.—*a.* Just as the individual letters constituting a large volume of plain text have more or less characteristic or fixed frequencies, so it is found that digraphs and trigraphs have characteristic frequencies, when a large volume of text is studied statistically. In Appendix 1, Table 6, are shown the relative frequencies of all digraphs appearing in the 260 telegrams referred to in Paragraph 9*e*. It will be noted that 428 of the 676 possible pairs of letters occur in these telegrams, but whereas many of them occur but once or twice, there are a few which occur hundreds of times.

b. In Appendix 1 will also be found several other kinds of tables and lists which will be useful to the student in his work, such as the relative order of frequency of the 50 digraphs of greatest

[1] This possible step is mentioned here for the purpose of making it clear that the plain-component sequence completion method cannot solve a case in which transposition has followed or preceded monoalphabetic substitution with standard alphabets. Cases of this kind will be discussed in a later text. It is sufficient to indicate at this point that the frequency distribution for such a combined substitution-transposition cipher would present the characteristics of a standard alphabet cipher—and yet the method of completing the plain-component sequence would fail to bring out any plain text.

frequency, the relative order of frequency of doubled letters, doubled vowels, doubled consonants, and so on. It is suggested that the student refer to this appendix now, to gain an idea of the data available for his future reference. Just how these data may be employed will become apparent very shortly.

26. Preparation of the work sheet.—*a.* The details to be considered in this paragraph may at first appear to be superfluous but long experience has proved that systematization of the work, and preparation of the data in the most utilizable, condensed form is most advisable, even if this seems to take considerable time. In the first place if it merely serves to avoid interruptions and irritations occasioned by failure to have the data in an instantly available form, it will pay by saving mental wear and tear. In the second place, especially in the case of complicated cryptograms, painstaking care in these details, while it may not always bring about success, is often the factor that is of greatest assistance in ultimate solution. The detailed preparation of the data may be irksome to the student, and he may be tempted to avoid as much of it as possible, but, unfortunately, in the early stages of solving a cryptogram he does not know (nor, for that matter, does the expert always know) just which data are essential and which may be neglected. Even though not all of the data may turn out to have been necessary, as a general rule, time is saved in the end if all the usual data are prepared as a regular preliminary to the solution of most cryptograms.

b. First, the cryptogram is recopied in the form of a *work sheet.* This sheet should be of a good quality of paper so as to withstand considerable erasure. If the cryptogram is to be copied by hand, cross-section paper of ¼-inch squares is extremely useful. The writing should be in ink, and plain, carefully made roman capital letters should be used in all cases. If the cryptogram is to be copied on a typewriter, the ribbon employed should be impregnated with an ink that will not smear or smudge under the hand.

c. The arrangement of the characters of the cryptogram on the work sheet is a matter of considerable importance. If the cryptogram as first obtained is in groups of regular length (usually five characters to a group) and if the uniliteral frequency distribution shows the cryptogram to be monoalphabetic, the characters should be copied without regard to this grouping. It is advisable to allow two spaces between letters, and to write a constant number of letters per line, approximately 25. At least two spaces, preferably three spaces, should be left between horizontal lines. Care should be taken to avoid crowding the letters in any case, for this is not only confusing to the eye but also mentally irritating when later it is found that not enough space has been left for making various sorts of marks or indications. If the cryptogram is originally in what appears to be word lengths (and this is the case, as a rule, only with the cryptograms of amateurs), naturally it should be copied on the work sheet in the original groupings. If further study of a cryptogram shows that some special grouping is required, it is often best to recopy it on a fresh work sheet rather than to attempt to indicate the new grouping on the old work sheet.

d. In order to be able to locate or refer to specific letters or groups of letters with speed, certainty, and without possibility of confusion, it is advisable to use coordinates applied to the lines and columns of the text as it appears on the work sheet. To minimize possibility of confusion, it is best to apply letters to the horizontal lines of the text, numbers to the vertical columns. In referring to a letter the horizontal line in which the letter is located is usually given first. Thus, referring to the work sheet shown below, coordinates A17 designate the letter Y, the 17th letter in the first line. The letter I is usually omitted from the series of line indicators so as to avoid confusion with the figure 1. If lines are limited to 25 letters each, then each set of 100 letters of the text is automatically blocked off by remembering that 4 lines constitute 100 letters.

e. Above each character of the cipher text may be some indication of the frequency of that character in the whole cryptogram. This indication may be the actual number of times the

character occurs, or, if colored pencils are used, the cipher letters may be divided up into three categories or groups—high frequency, medium frequency, and low frequency. It is perhaps simpler, if clerical help is available, to indicate the actual frequencies. This saves constant reference to the frequency tables, which interrupts the train of thought, and saves considerable time in the end.

f. After the special frequency distribution, explained in Par. 27 below, has been constructed, repetitions of digraphs and trigraphs should be underscored. In so doing, the student should be particularly watchful of trigraphic repetitions which can be further extended into tetragraphs and polygraphs of greater length. Repetitions of more than ten characters should be set off by heavy vertical lines, as they indicate repeated phrases and are of considerable assistance in solution. If a repetition continues from one line to the next, put an arrow at the end of the underscore to signal this fact. Reversible digraphs should also be indicated by an underscore with an arrow pointing in both directions. Anything which strikes the eye as being peculiar, unusual, or significant as regards the distribution or recurrence of the characters should be noted. All these marks should, if convenient, be made with ink so as not to cause smudging. The work sheet will now appear as shown herewith (not all the repetitions are underscored):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	10	19	23	35	19	10	3	19	15	5	5	35	19	19	35	23	14	10	5	19	15	4	35	23	10
A	S	F	D	Z	F	I	O	G	H	L	P	Z	F	G	Z	D	Y	S	P	F	H	B	Z	D	S
	19	16	15	22	19	5	5	5	16	23	19	19	14	16	3	16	19	16	15	22	19	3	23	35	35
B	G	V	H	T	F	U	P	L	V	D	F	G	Y	V	J	V	F	V	H	T	G	A	D	Z	Z
	3	10	22	14	23	35	14	19	35	3	35	22	19	5	22	16	22	35	4	23	16	19	15	22	35
C	A	I	T	Y	D	Z	Y	F	Z	J	Z	T	G	P	T	V	T	Z	B	D	V	F	H	T	Z
	23	19	3	10	4	19	10	23	35	14	16	22	3	3	10	14	16	22	3	19	16	2	19	35	35
D	D	F	X	S	B	G	I	D	Z	Y	V	T	X	O	I	Y	V	T	E	F	V	M	G	Z	Z
	22	15	5	5	16	3	35	23	19	2	15	22	35	3	10	22	14	23	35	14	4	23	16	19	15
E	T	H	L	L	V	X	Z	D	F	M	H	T	Z	A	I	T	Y	D	Z	Y	B	D	V	F	H
	22	35	23	19	2	35	23	35	35	3	10	3	10	10	19	35	14	19	3	16	19	10	5	19	35
F	T	Z	D	F	K	Z	D	Z	Z	J	S	X	I	S	G	Z	Y	G	A	V	F	S	L	G	Z
	23	22	15	15	22	1	23	35	2	10	16	22	14	35	23	3	35	19	19	15	22	35	3	10	22
G	D	T	H	H	T	C	D	Z	R	S	V	T	Y	Z	D	O	Z	F	F	H	T	Z	A	I	T
	14	23	35	14	19	3	16	23	19	35	35	22	2	15	10	22	14	35	14	10	23	35	19	15	5
H	Y	D	Z	Y	G	A	V	D	G	Z	Z	T	K	H	I	T	Y	Z	Y	S	D	Z	G	H	U
	35	19	35	22	19	5	5	19	23	10	3	1	19	15	3	3	10	2	5	35	23	19	5	10	23
J	Z	F	Z	T	G	U	P	G	D	I	X	W	G	H	X	A	S	R	U	Z	D	F	U	I	D
	3	19	15	22	16	3	3	19	3	3															
K	E	G	H	T	V	E	A	G	X	X															

27. Triliteral-frequency distributions.—*a.* In what has gone before, a type of frequency distribution known as a uniliteral frequency distribution was used. This, of course, shows only the number of times each individual letter occurs. In order to apply the normal digraphic and

trigraphic frequency data (given in Appendix 1) to the solution of a cryptogram of the type now being studied, it is obvious that the data with respect to digraphs and trigraphs occurring in the cryptogram should be compiled and should be compared with the data for normal plain text. In order to accomplish this in suitable manner, it is advisable to construct a slightly more complicated form of distribution termed a *triliteral frequency distribution.*[2]

b. Given a cryptogram of 50 or more letters and the task of determining what trigraphs are present in the cryptogram, there are three ways in which the data may be arranged or assembled. One may require that the data show (1) each letter with its two succeeding letters; (2) each letter with its two preceding letters; (3) each letter with one preceding letter and one succeeding letter.

c. A distribution of the first of the three foregoing types may be designated as a "triliteral frequency distribution showing two suffixes"; the second type may be designated as a "triliteral frequency distribution showing two prefixes"; the third type may be designated as a "triliteral frequency distribution showing one prefix and one suffix." Quadriliteral and pentaliteral frequency distributions may occasionally be found useful.

d. Which of these three arrangements is to be employed at a specific time depends largely upon what the data are intended to show. For present purposes, in connection with the solution of a monoalphabetic substitution cipher employing a mixed alphabet, possibly the third arrangement, that showing one prefix and one suffix, is most satisfactory.

e. It is convenient to use ¼-inch cross-section paper for the construction of a triliteral frequency distribution in the form of a distribution showing crests and troughs, such as that in Figure 14. In that figure the prefix to each letter to be recorded is inserted in the left half of the cell directly above the cipher letter being recorded; the suffix to each letter is inserted in the right half of the cell directly above the letter being recorded; and in each case the prefix and the suffix to the letter being recorded occupy the same cell, the prefix being directly to the left of the suffix. The number in parentheses gives the total frequency for each letter.

---

[2] Heretofore such a distribution has been termed a "trigraphic frequency table." It is thought that the word "triliteral" is more suitable, to correspond with the designation "uniliteral" in the case of the distribution of the single letters. A trigraphic distribution of A B C D E F would consider only the trigraphs A B C and D E F, whereas a triliteral distribution would consider the groups A B C, B C D, C D E, and D E F. (See also Par. 11*d*.) The use of the word "distribution" to replace the word "table" has already been explained.

CONDENSED TABLE OF REPETITIONS

Digraphs			*Trigraphs*		*Longer Polygraphs*
DZ–9	TZ–5	VF–4	DZY–4	FHT–3	HTZAITYDZY–2
ZD–9	TY–5	VT–4	HTZ–4	TYD–3	BDVFHTZDF–2
HT–8	FH–4	ZF–4	ITY–4	YDZ–3	ZAITYDZY–3
ZY–6	GH–4	ZT–4	ZDF–4	ZAI–3	FHTZ–3
DF–5	IT–4	ZZ–4	AIT–3		
GZ–5					

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
(8)	(4)	(1)	(28)	(3)	(29)	(19)	(16)	(10)	(8)	(2)	(5)	(3)	(0)	(2)	(5)	(0)	(2)	(16)	(22)	(6)	(16)	(1)	(8)	(14)	(36)
EG			IE																HV					UD	
XS			ZF																ZG					FT	
GV			GI																IY					UF	
ZI			SZ																ZK					DG	
GV	YD		VG		DU	AX													IY					YY	
ZI	SG		YZ		ZZ	EH													HZ					ZT	
ZI	ZD		ZO		FH	WH													VY			TE		GZ	
GD	HZ	TD	CZ	DG	ZF	PD													HC			AD		DY	
			ZT	TF	VS	TU	GT												DH			ST		TA	
			ZZ	SD	DK	ZH	GX												HZ			AF		XD	
			ZF	OH	VH	DZ	GU												IY			DF		ZS	ZT
			BV	GL	DM	YA	KI												HZ			LX		TZ	GZ
			YZ	FO	EV	LZ	FT												HZ			FM		ZG	DY
			ZF	VV	DX	YA	HT	UD										AR	ZH			YT		TD	TD
			IZ	FZ	VH	SZ	FH	DX										YD	VE			YT	X–	TZ	TB
EG			ZF	HP	YZ	MZ	TT	HT										RV	VX			YT	GX	ZG	JT
XS			BV	VG	VV	BI	MT	AT										FL	HZ			DF	HA	ZB	FJ
GV			YZ	IG	DG	TP	TL	XS										IG	VZ			TT	IW	IV	ZA
ZI			AZ	LZ	TU	TA	FT	AT				SG				UG		JX	PV	FI	FH		ZV	ZV	DZ
GV	YD		VF		PH	FY	VT	OY				LV				GT		XB	ZG	RZ	JF		SI	ZF	BD
ZI	SG		ZS	ZS	ZG	SV	VT	GD	ZS			HL			DZ	UL		DG	IY	GP	YJ		VZ	TD	GD
ZI	ZD		ZY	DG	ZI	FZ	FB	AT	ZZ	TH	PV	FH			XI	SF	SU	YP	HG	HZ	LD		TO	GV	PF
GD	HZ	TD	FZ	TF	SD	OH	GL	FO	VV	FZ	HP	VG			IG	LZ	ZS	–F	HF	FP	GH	XG	FS	DS	DF

FIGURE 14.

f. The triliteral frequency distribution is now to be examined with a view to ascertaining what digraphs and trigraphs occur two or more times in the cryptogram. Consider the pair of columns containing the prefixes and suffixes to D_c in the distribution, as shown in Fig. 14. This pair of columns shows that the following digraphs appear in the cryptogram:

Digraphs based on prefixes (arranged as one reads up the column)	*Digraphs based on suffixes (arranged as one reads up the column)*
FD, ZD, ZD, VD, AD, YD, BD,	DZ, DY, DS, DF, DZ, DZ, DV,
ZD, ID, ZD, YD, BD, ZD, ZD,	DF, DZ, DF, DZ, DV, DF, DZ,
ZD, CD, ZD, YD, VD, SD, GD,	DT, DZ, DO, DZ, DG, DZ, DI,
ZD, ID	DF, DE

The nature of the triliteral frequency distribution is such that in finding what digraphs are present in the cryptogram it is immaterial whether the prefixes or the suffixes to the cipher letters are studied, *so long as one is consistent in the study.* For example, in the foregoing list of digraphs based on the prefixes to D_c, the digraphs FD, ZD, ZD, VD, etc., are found; if now, the student will refer to the suffixes of F_c, Z_c, V_c, etc., he will find the very same digraphs indicated. This being the case, the question may be raised as to what value there is in listing both the prefixes and the suffixes to the cipher letters. The answer is that by so doing the trigraphs are indicated at the same time. For example, in the case of D_c, the following trigraphs are indicated:

FDZ, ZDY, ZDS, VDF, ADZ, YDZ, BDV, ZDF, IDZ, ZDF, YDZ, BDV, ZDF,
ZDZ, ZDT, CDZ, ZDO, YDZ, VDG, SDZ, GDI, ZDF, IDE.

g. The *repeated* digraphs and trigraphs can now be found quite readily. Thus, in the case of D_c, examining the list of digraphs based on suffixes, the following repetitions are noted:

DZ appears 9 times
DF appears 5 times
DV appears 2 times

Examining the trigraphs with D_c as central letter, the following repetitions are noted:

ZDF appears 4 times
YDZ appears 3 times
BDV appears 2 times

h. It is unnecessary, of course, to go through the detailed procedure set forth in the preceding subparagraphs in order to find all the repeated digraphs and trigraphs. The repeated trigraphs with D_c as central letter can be found merely from an inspection of the prefixes and suffixes opposite D_c in the distribution. It is necessary only to find those cases in which two or more prefixes are identical at the same time that the suffixes are identical. For example, the distribution shows at once that in four cases the prefix to D_c is Z_c at the same time that the suffix to this letter is F_c. Hence, the trigraph ZDF appears four times. The repeated trigraphs may all be found in this manner.

i. The most frequently repeated digraphs and trigraphs are then assembled in what is termed a *condensed table of repetitions*, so as to bring this information prominently before the eye. As a rule, digraphs which occur less than four or five times, and trigraphs which occur less than three or four times may be omitted from the condensed table as being relatively of no importance in the study of repetitions. In the condensed table the frequencies of the individual letters forming the most important digraphs, trigraphs, etc., should be indicated.

28. Classifying the cipher letters into vowels and consonants.—*a.* Before proceeding to a detailed analysis of the repeated digraphs and trigraphs, a very important step can be taken which will be of assistance not only in the analysis of the repetitions but also in the final solution of the cryptogram. This step concerns the classification of the high-frequency letters into two

groups—vowels and consonants. For if the cryptanalyst can quickly ascertain the equivalents of the four vowels, A, E, I, and O, and of only the four consonants, N, R, S, and T, he will then have the values of approximately two-thirds of all the cipher letters that occur in the cryptogram; the values of the remaining letters can almost be filled in automatically.

b. The basis for the classification will be found to rest upon a comparatively simple phenomenon: the associational or combinatory behavior of vowels is, in general, quite different from that of consonants. If an examination be made of Table 7–B in Appendix 1, showing the relative order of frequency of the 18 digraphs composing 25 percent of English telegraphic text, it will be seen that the letter E enters into the composition of 9 of the 18 digraphs; that is, in exactly half of all the cases the letter E is one of the two letters forming the digraph. The digraphs containing E are as follows:

ED EN ER ES
NE RE SE TE VE

The remaining nine digraphs are as follows:

AN ND OR ST
IN NT TH
ON TO

c. None of the 18 digraphs is a combination of vowels. Note now that of the 9 combinations with E, 7 are with the consonants N, R, S, and T, one is with D, one is with V, and *none is with any vowel.* In other words, E_p combines most readily with consonants but not with other vowels, or even with itself. Using the terms often employed in the chemical analogy, E shows a great "affinity" for the consonants N, R, S, T, but not for the vowels. Therefore, if the letters of highest frequency occurring in a given cryptogram are listed, together with the number of times each of them combines with the cipher equivalent of E_p, those which show considerable combining power or affinity for the cipher equivalent of E_p may be assumed to be the cipher equivalents of N, R, S, T_p; those which do not show any affinity for the cipher equivalent of E_p may be assumed to be the cipher equivalents of A, I, O, U_p. Applying these principles to the problem in hand, and examining the triliteral frequency distribution, it is quite certain that $Z_c = E_p$, not only because Z_c is the letter of highest frequency, but also because it combines with *several* other high-frequency letters, such as D_c, F_c, G_c, etc. The nine letters of next highest frequency are:

23	22	19	19	16	15	14	10	10
D	T	F	G	V	H	Y	S	I

Let the combinations these letters form with Z_c be indicated in the following manner:

Number of times Z_c occurs as prefix	𝍸	𝍷	𝍷	–			𝍸	
Cipher Letter	D(23)	T(22)	F(19)	G(19)	V(16)	H(15)	Y(14)	S(10) I(10)
Number of times Z_c occurs as suffix	𝍸	𝍸	=	𝍸			=	

d. Consider D_c. It occurs 23 times in the message and 18 of those times it is combined with Z_c, 9 times in the form Z_cD_c ($=E\Theta_p$), and 9 times in the form D_cZ_c ($=\Theta E_p$). It is clear that D_c must be a consonant. In the same way, consider T_c, which shows 9 combinations with Z_c, 4 in the form Z_cT_c ($=E\Theta_p$) and 5 in the form T_cZ_c ($=\Theta E_p$). The letter T_c appears to represent a consonant, as do also the letters F_c, G_c, and Y_c. On the other hand, consider V_c, occurring in all 16 times but never in combination with Z_c; it appears to represent a vowel, as do also the letters H_c, S_c, and I_c. So far, then, the following classification would seem logical:

Vowels	*Consonants*
$Z_c(=E_p)$, V_c, H_c, S_c, I_c	D_c, T_c, F_c, G_c, Y_c

29. Further analysis of the letters representing vowels and consonants.—*a.* 0_p is usually the vowel of second highest frequency. Is it possible to determine which of the letters V, H, S, I_c is the cipher equivalent of 0_p? Let reference be made again to Table 6 in Appendix 1, where it is seen that the 10 most frequently occurring diphthongs are:

Diphthong	IO	OU	EA	EI	AI	IE	AU	EO	AY	UE
Frequency	41	37	35	27	17	13	13	12	12	11

If V, H, S, I_c are really the cipher equivalents of A, I, O, U_p (not respectively), perhaps it is possible to determine which is which *by examining the combinations they make among themselves and with* Z_c ($=E_p$). Let the combinations of V, H, S, I, and Z that occur in the message be listed. There are only the following:

ZZ_c—4	HI—1
VH —2	SV—1
HH —1	IS—1

ZZ_c is of course EE_p. Note the doublet HH_c; if H_c is a vowel, then the chances are excellent that $H_c=0_p$ because the doublets AA_p, II_p, UU_p, are practically non-existent, whereas the double vowel combination 00_p is of next highest frequency to the double vowel combination EE_p. If $H_c=0_p$, then V_c must be I_p because the digraph VH_c occurring two times in the message could hardly be $A0_p$, or $U0_p$, whereas the diphthong $I0_p$ is the one of high frequency in English. So far then, the tentative (because so far unverified) results of the analysis are as follows:

$$Z_c=E_p \quad H_c=0_p \quad V_c=I_p$$

This leaves only two letters, I_c and S_c (already classified as vowels) to be separated into A_p and U_p. Note the digraphs:

$$HI_c=0\theta_p$$
$$SV_c=\theta I_p$$
$$IS_c=\theta\theta_p$$

Only two alternatives are open:

(1) Either $I_c=A_p$ and $S_c=U_p$,
(2) Or $I_c=U_p$ and $S_c=A_p$.

If the first alternative is selected, then

$$HI_c=0A_p$$
$$SV_c=UI_p$$
$$IS_c=AU_p$$

If the second alternative is selected, then

$$HI_c=0U_p$$
$$SV_c=AI_p$$
$$IS_c=UA_p$$

The eye finds it difficult to choose between these alternatives; but suppose the frequency values of the plain-text diphthongs as given in Table 6 of Appendix 1 are added for each of these alternatives, giving the following:

$HI_c=0A_p$, frequency value= 7	$HI_c=0U_p$, frequency value=37
$SV_c=UI_p$, frequency value= 5	$SV_c=AI_p$, frequency value=17
$IS_c=AU_p$, frequency value=13	$IS_c=UA_p$, frequency value= 5
Total..... 25	Total..... 59

Mathematically, the second alternative is more than twice as probable as the first. Let it be assumed to be correct and the following (still tentative) values are now at hand:

$$Z_c=E_p \quad H_c=O_p \quad V_c=I_p \quad S_c=A_p \quad I_c=U_p$$

b. Attention is now directed to the letters classified as consonants. How far is it possible to ascertain their values? The letter D_c, from considerations of frequency alone, would seem to be T_p, but its frequency, 23, is not considerably greater than that for T_c. It is not much greater than that for F_c or G_c, with a frequency of 19 each. But perhaps it is possible to ascertain not the value of one letter alone but of two letters at one stroke. To do this one may make use of a tetragraph of considerable importance in English, *viz*, $TION_p$. For if the analysis pertaining to the vowels is correct, and if $VH_c=IO_p$, then an examination of the letters immediately before and after the digraph VH_c in the cipher text might disclose both T_p and N_p. Reference to the text gives the following:

$$\begin{array}{ll} GVHT_c & FVHT_c \\ \theta IO\theta_p & \theta IO\theta_p \end{array}$$

The letter T_c follows VH_c in both cases and very probably indicates that $T_c=N_p$; but as to whether G_c or F_c equals T_p cannot be decided. However, two conclusions are clear: first, the letter D_c is neither T_p nor N_p, from which it follows that it must be either R_p or S_p; second, the letters G_c and F_c must be either T_p and S_p, respectively, or S_p and T_p, respectively, because the only tetragraphs usually found (in English) containing the diphthong IO_p as central letters are $SION_p$ and $TION_p$. This in turn means that as regards D_c, the latter cannot be *either* R_p or S_p; it *must* be R_p, a conclusion which is corroborated by the fact that ZD_c (=ER_p) and DZ_c (=RE_p) occur 9 times each. Thus far, then, the identifications, when inserted in an *enciphering* alphabet, are as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S				Z				V					T	H			D	G	F	I					
																			F	G						

30. Substituting deduced values in the cryptogram.—*a.* Thus far the analysis has been almost purely hypothetical, for as yet not a single one of the values deduced from the foregoing analysis has been tried out in the cryptogram. It is high time that this be done, because the final test of the validity of the hypotheses, assumptions, and identifications made in any cryptographic study is, after all, only this: do these hypotheses, assumptions, and identifications ultimately yield verifiable, intelligible plain-text when *consistently* applied to the cipher text?

b. At the present stage in the process, since there are at hand the assumed values of but 9 out of the 25 letters that appear, it is obvious that a continuous "reading" of the cryptogram can certainly not be expected from a mere insertion of the values of the 9 letters. However, the substitution of these values should do two things. First, it should immediately disclose the fragments, outlines, or "skeletons" of "good" words in the text; and second, it should disclose no places in the text where "impossible" sequences of letters are established. By the first is meant that the partially deciphered text should show the outlines or skeletons of words such as may be expected to be found in the communication; this will become quite clear in the next subparagraph. By the second is meant that sequences, such as "AOOEN" or "TNRSENO" or the like, obviously not possible or extremely unusual in normal English text, must not result from the substitution of the tentative identifications resulting from the analysis. The appearance of several such extremely unusual or impossible sequences at once signifies that one or more of the assumed values is incorrect.

c. Here are the results of substituting the nine values which have been deduced by the reasoning based on a classification of the high-frequency letters into vowels and consonants and the study of the members of the two groups:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	10 S A	19 F T S	23 D R	35 Z E	19 F T S	10 I	3 O	19 G S T	15 H O	5 L	5 P	35 Z E	19 F T S	19 G S T	35 Z E	23 D R	14 Y	10 S A	5 P	19 F T S	15 H O	4 B	35 Z E	23 D R	10 S A
B	19 G S T	16 V I	15 H O	22 T N	19 F T S	8 U	5 P	5 L	16 V I	23 D R	19 F T S	19 G S T	14 Y	16 V I	3 J	16 V I	19 F T S	16 V I	15 H O	22 T N	19 G S T	8 A	23 D R	35 Z E	35 Z E
C	8 A	10 I	22 T N	14 Y	23 D R	35 Z E	14 Y	19 F T S	35 Z E	3 J	35 Z E	22 T N	19 G S T	5 P	22 T N	16 V I	22 T N	35 Z E	4 B	23 D R	16 V I	19 F T S	15 H O	22 T N	35 Z E
D	23 D R	19 F T S	8 X	10 S A	4 B	19 G S T	10 I	23 D R	35 Z E	14 Y	16 V I	22 T N	8 X	3 O	10 I	14 Y	16 V I	22 T N	3 E	19 F T S	16 V I	2 M	19 G S T	35 Z E	35 Z E
E	22 T N	15 H O	5 L	5 L	16 V I	8 X	35 Z E	23 D R	19 F T S	2 M	15 H O	23 T N	35 Z E	8 A	10 I	22 T N	14 Y	23 D R	35 Z E	14 Y	4 B	23 D R	16 V I	19 F T S	15 H O
F	22 T N	35 Z E	23 D R	19 F T S	2 K	35 Z E	23 D R	35 Z E	35 Z E	3 J	10 S A	8 X	10 I	10 S A	19 G S T	35 Z E	14 Y	19 G S T	8 A	16 V I	19 F T S	10 S A	5 L	19 G S T	35 Z E
G	23 D R	22 T N	15 H O	15 H O	22 T N	1 C	23 D R	35 Z E	2 R	10 S A	16 V I	22 T N	14 Y	35 Z E	23 D R	3 O	35 Z E	19 F T S	19 F T S	15 H O	22 T N	35 Z E	8 A	10 I	22 T N
H	14 Y	23 D R	35 Z E	14 Y	19 G S T	8 A	16 V I	23 D R	19 G S T	35 Z E	35 Z E	22 T N	2 K	15 H O	10 I	22 T N	14 Y	35 Z E	14 Y	10 S A	23 D R	35 Z E	19 G S T	15 H O	5 U
J	35 Z E	19 F T S	35 Z E	22 T N	19 G S T	5 U	5 P	19 G S T	23 D R	10 I	8 X	1 W	24 G S T	15 H O	8 X	8 A	10 S A	2 R	5 U	35 Z E	23 D R	19 F T S	5 U	10 I	23 D R
K	8 E	19 G S T	15 H O	22 T N	16 V I	8 E	8 A	19 G S T	8 X	8 X															

d. No impossible sequences are brought to light, and, moreover, several long words, nearly complete, stand out in the text. Note the following portions:

A21
H B Z D S G V H T F
(1) O ? E R A S I O N T
T S

C15
T V T Z B D V F H T Z D F
(2) N I N E ? R I T O N E R T
S S

F22
S L G Z D T H H T
(3) A ? S E R N O O N
T

The words are obviously OPERATIONS, NINE PRISONERS, and AFTERNOON. The value G_c si clearly T_p; that of F_c is S_p; and the following additional values are certain:

$$B_c=P_p \qquad L_c=F_p$$

31. Completing the solution.—*a.* Each time an additional value is obtained, substitution is at once made throughout the cryptogram. This leads to the determination of further values, in an ever-widening circle, until all the identifications are firmly and finally established, and the message is completely solved. In this case the decipherment is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	S	F	D	Z	F	I	O	G	H	L	P	Z	F	G	Z	D	Y	S	P	F	H	B	Z	D	S
	A	S	R	E	S	U	L	T	O	F	Y	E	S	T	E	R	D	A	Y	S	O	P	E	R	A
B	G	V	H	T	F	U	P	L	V	D	F	G	Y	V	J	V	F	V	H	T	G	A	D	Z	Z
	T	I	O	N	S	B	Y	F	I	R	S	T	D	I	V	I	S	I	O	N	T	H	R	E	E
C	A	I	T	Y	D	Z	Y	F	Z	J	Z	T	G	P	T	V	T	Z	B	D	V	F	H	T	Z
	H	U	N	D	R	E	D	S	E	V	E	N	T	Y	N	I	N	E	P	R	I	S	O	N	E
D	D	F	X	S	B	G	I	D	Z	Y	V	T	X	O	I	Y	V	T	E	F	V	M	G	Z	Z
	R	S	C	A	P	T	U	R	E	D	I	N	C	L	U	D	I	N	G	S	I	X	T	E	E
E	T	H	L	L	V	X	Z	D	F	M	H	T	Z	A	I	T	Y	D	Z	Y	B	D	V	F	H
	N	O	F	F	I	C	E	R	S	X	O	N	E	H	U	N	D	R	E	D	P	R	I	S	O
F	T	Z	D	F	K	Z	D	Z	Z	J	S	X	I	S	G	Z	Y	G	A	V	F	S	L	G	Z
	N	E	R	S	W	E	R	E	E	V	A	C	U	A	T	E	D	T	H	I	S	A	F	T	E
G	D	T	H	H	T	C	D	Z	R	S	V	T	Y	Z	D	O	Z	F	F	H	T	Z	A	I	T
	R	N	O	O	N	Q	R	E	M	A	I	N	D	E	R	L	E	S	S	O	N	E	H	U	N
H	Y	D	Z	Y	G	A	V	D	G	Z	Z	T	K	H	I	T	Y	Z	Y	S	D	Z	G	H	U
	D	R	E	D	T	H	I	R	T	E	E	N	W	O	U	N	D	E	D	A	R	E	T	O	B
J	Z	F	Z	T	G	U	P	G	D	I	X	W	G	H	X	A	S	R	U	Z	D	F	U	I	D
	E	S	E	N	T	B	Y	T	R	U	C	K	T	O	C	H	A	M	B	E	R	S	B	U	R
K	E	G	H	T	V	E	A	G	X	X															
	G	T	O	N	I	G	H	T	X	X															

Message: AS RESULT OF YESTERDAYS OPERATIONS BY FIRST DIVISION THREE HUNDRED SEVENTY NINE PRISONERS CAPTURED INCLUDING SIXTEEN OFFICERS ONE HUNDRED PRISONERS WERE EVACUATED THIS AFTERNOON REMAINDER LESS ONE HUNDRED THIRTEEN WOUNDED ARE TO BE SENT BY TRUCK TO CHAMBERSBURG TONIGHT

b. The solution should, as a rule, not be considered complete until an attempt has been made to discover all the elements underlying the general system and the specific key to a message. In this case, there is no need to delve further into the general system, for it is merely one of monoalphabetic substitution with a mixed cipher alphabet. It is necessary or advisable, however, to reconstruct the cipher alphabet because this may give clues that later may become valuable.

c. Cipher alphabets should, as a rule, be reconstructed by the cryptanalyst in the form of *enciphering* alphabets because they will then usually be in the form in which the encipherer used them. This is important for two reasons. First, if the sequence in the cipher component gives evidence of system in its construction or if it yields clues pointing toward its derivation from a keyword or a key-phrase, this may often corroborate the identifications already made and may lead directly to additional identifications. A word or two of explanation is advisable here. For example, refer to the skeletonized enciphering alphabet given at the end of par. 29*b*:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S				Z				V					T	H			D	G	F	I					
																		F	G							

Suppose the cryptanalyst, looking at the sequence DGFI or DFGI in the cipher component, suspects the presence of a keyword-mixed alphabet. Then DFGI is certainly a more plausible sequence than DGFI. Again, noting the sequence S . . . Z . . . V . . . . TH . . D, he might have an idea that the keyword begins after the Z and that the TH is followed by AB or BC. This would mean that either P, Q_p=A, B_c or B, C_c. Assuming that P, Q_p=A, B_c, he refers to the frequency distribution and finds that the assumptions P_p=A_c and Q_p=B_c are not good; on the other hand, assuming that P, Q_p=B, C_c, the frequency distribution gives excellent corroboration. A trial of these values would materially hasten solution because it is often the case in cryptanalysis that if the value of a very low-frequency letter can be surely established it will yield clues to other values very quickly. Thus, if Q_p is definitely identified it almost invariably will identify U_p, and will give clues to the letter following the U_p, since it must be a vowel. In the case under discussion the identification PQ_p=BC_c would have turned out to be correct. For the foregoing reason an attempt should always be made in the early stages of the analysis to determine, if possible, the basis of construction or derivation of the cipher alphabet; as a rule this can be done only by means of the enciphering alphabet, and not the deciphering alphabet. For example, the skeletonized *deciphering* alphabet corresponding to the enciphering alphabet directly above is as follows:

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain					R	T	S	O	U										A	N		I				E
						S	T																			

Here no evidences of a keyword-mixed alphabet are seen at all. However, if the enciphering alphabet has been examined and shows no evidences of systematic construction, the deciphering alphabet should then be examined with this in view, because occasionally it is the deciphering alphabet which shows the presence of a key or keying element, or which has been systematically derived from a word or phrase. The second reason why it is important to try to discover the basis

of construction or derivation of the cipher alphabet is that it affords clues to the general type of keywords or keying elements employed by the enemy. This is a psychological factor, of course, and may be of assistance in subsequent studies of his traffic. It merely gives a clue to the general type of thinking indulged in by certain of his cryptographers.

d. In the case of the foregoing solution, the complete enciphering alphabet is found to be as follows:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	

Obviously, the letter Q, which is the only letter not appearing in the cryptogram, should follow P in the cipher component. Note now that the latter is based upon the keyword LEAVENWORTH, and that this particular cipher alphabet has been composed by shifting the mixed sequence based upon this keyword five intervals to the right so that the key for the message is $A_p=S_c$. Note also that the deciphering alphabet fails to give any evidence of keyword construction based upon the word LEAVENWORTH.

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	H	P	Q	R	G	S	T	O	U	V	W	F	X	J	L	Y	Z	M	A	N	B	I	K	C	D	E

e. If neither the enciphering or the deciphering alphabet exhibits characteristics which give indication of derivation from a keyword by some form of mixing or disarrangement, the latter is nevertheless not finally excluded as a possibility. The student is referred to Section IX of *Elementary Military Cryptography*, wherein will be found methods for deriving mixed alphabets by transposition methods applied to keyword-mixed alphabets. For the reconstruction of such mixed alphabets the cryptanalyst must use ingenuity and a knowledge of the more common methods of suppressing the appearance of keywords in the mixed alphabets.

32. General notes on the foregoing solution.—*a.* The example solved above is admittedly a more or less artificial illustration of the steps in analysis, made so in order to demonstrate general principles. It was easy to solve because the frequencies of the various cipher letters corresponded quite well with the normal or expected frequencies. However, all cryptograms of the same monoalphabetical nature can be solved along the same general lines, after more or less experimentation, depending upon the length of the cryptogram, the skill, and the experience of the cryptanalyst.

b. It is no cause for discouragement if the student's initial attempts to solve a cryptogram of this type require much more time and effort than were apparently required in solving the foregoing purely illustrative example. It is indeed rarely the case that *every* assumption made by the cryptanalyst proves in the end to have been correct; more often is it the case that a good many of his initial assumptions are incorrect, and that he loses much time in casting out the erroneous ones. The speed and facility with which this elimination process is conducted is in many cases all that distinguishes the expert from the novice.

c. Nor will the student always find that the initial classification into vowels and consonants can be accomplished as easily and quickly as was apparently the case in the illustrative example. The principles indicated are very general in their nature and applicability, and there are, in addition, some other principles that may be brought to bear in case of difficulty. Of these, perhaps the most useful are the following:

(1) In normal English it is unusual to find two or three consonants in succession, each of high frequency. If in a cryptogram a succession of three or four letters of high-frequency appear in succession, it is practically certain that at least one of these represents a vowel.[3]

[3] Sequences of seven consonants are not impossible, however, as in STRENGTH THROUGH.

(2) Successions of three vowels are rather unusual in English.[4] Practically the only time this happens is when a word ends in two vowels and the next word begins with a vowel.[5]

(3) When two letters already classified as vowel-equivalents are separated by a sequence of six or more letters, it is either the case that one of the supposed vowel-equivalents is incorrect, or else that one or more of the intermediate letters is a vowel-equivalent.[6]

(4) Reference to Table 7–B of Appendix 1 discloses the following:

Distribution of first 18 digraphs forming 25 percent of English text

Number of consonant-consonant digraphs	4
Number of consonant-vowel digraphs	6
Number of vowel-consonant digraphs	8
Number of vowel-vowel digraphs	0

Distribution of first 53 digraphs forming 50 percent of English text

Number of consonant-consonant digraphs	8
Number of consonant-vowel digraphs	23
Number of vowel-consonant digraphs	18
Number of vowel-vowel digraphs	4

The latter tabulation shows that of the first 53 digraphs which form 50 percent of English text, 41 of them, that is, over 75 percent, are combinations of a vowel with a consonant. In short, in normal English the vowels and the high-frequency consonants are in the long run distributed fairly evenly and regularly throughout the text.

(5) As a rule, repetitions of trigraphs in the cipher text are composed of high-frequency letters forming high-frequency combinations. The latter practically always contain at least one vowel; in fact, if reference is made to Table 10–A of Appendix 1, it will be noted that 36 of the 56 trigraphs having a frequency of 100 or more contain one vowel, 17 of them contain two vowels, and only three of them contain no vowel. In the case of tetragraph repetitions, Table 11–A of Appendix 1 shows that no tetragraph listed therein fails to contain at least one vowel; 28 of them contain one vowel, 25 contain two vowels, and 2 contain three vowels.

(6) Quite frequently when two known vowel-equivalents are separated by six or more letters none of which seems to be of sufficiently high frequency to represent one of the vowels A E I O, the chances are good that the cipher-equivalent of the vowel U or Y is present.

(7) The letter Q is invariably followed by U; the letters J and V are invariably followed by a vowel.

d. In the foregoing example the amount of experimentation or "cutting and fitting" was practically nil. (This is not true of real cases as a rule.) Where such experimentation is neces-

---

[4] Note that the word RADIOED, past tense of the verb RADIO, is coming into usage.

[5] A sequence of seven vowels is not impossible, however, as in THE WAY YOU EARN.

[6] Some cryptanalysts place a good deal of emphasis upon this principle as a method of locating the remaining vowels after the first two or three have been located. They recommend that the latter be underlined throughout the text and then all sequences of five or more letters showing no underlines be studied attentively. Certain letters which occur in several such sequences are sure to be vowels. An arithmetical aid in the study is as follows: Take a letter thought to be a good possibility as the cipher equivalent of a vowel (hereafter termed a *possible vowel-equivalent*) and find the length of each interval from the possible vowel-equivalent to the next *known* (fairly surely determined) vowel-equivalent. Multiply the interval by the number of times this interval is found. Add the products and divide by the total number of intervals considered. This will give the *mean* interval for that possible vowel-equivalent. Do the same for all the other possible vowel-equivalents. The one for which the mean is the greatest is most probably a vowel-equivalent. Underline this letter throughout the text and repeat the process for locating additional vowel-equivalents, if any remain to be located.

sary, the underscoring of all repetitions of several letters is very essential, as it calls attention to peculiarities of structure that often yield clues.

e. After a few basic assumptions of values have been made, if short words or skeletons of words do not become manifest, it is necessary to make further assumptions for unidentified letters. This is accomplished most often by assuming a word.[7] Now there are two places in every message which lend themselves more readily to successful attack by the assumption of words than do any other places—the very beginning and the very end of the message. The reason is quite obvious, for although words may begin or end with almost any letter of the alphabet, they usually begin and end with but a few very common digraphs and trigraphs. Very often the association of letters in peculiar combinations will enable the student to note where one word ends and the next begins. For example suppose, E, N, S, and T have been definitely identified, and a sequence like the following is found in a cryptogram:

. . . E N T S N E . . .

Obviously the break between two words should fall either after the S of E N T S or after the T of E N T, so that two possibilities are offered: . . . E N T S / N E . . ., or . . . E N T / S N E . . .. Since in English there are very few words with the initial trigraph S N E, it is most likely that the proper division is . . . E N T S / N E . . .. Obviously, when several word divisions have been found, the solution is more readily achieved because of the greater ease with which assumptions of additional new values may be made.

33. The "probable word" method; its value and applicability.—*a.* In practically all cryptanalytic studies, short-cuts can often be made by assuming the presence of certain words in the message under study. Some writers attach so much value to this kind of an "attack from the rear" that they practically elevate it to the position of a method and call it the "intuitive method" or the "probable-word method." It is, of course, merely a refinement of what in every-day language is called "assuming" or "guessing" a word in the message. The value of making a "good guess" can hardly be overestimated, and the cryptanalyst should never feel that he is accomplishing a solution by an illegitimate subterfuge when he has made a fortunate guess leading to solution. A correct assumption as to plain text will often save hours or days of labor, and sometimes there is no alternative but to try to "guess a word", for occasionally a system is encountered the solution of which is absolutely dependent upon this artifice.

b. The expression "good guess" is used advisedly. For it is "good" in two respects. First, the cryptanalyst must use care in making his assumptions as to plain-text words. In this he must be guided by extraneous circumstances leading to the assumption of *probable* words—not just any words that come to his mind. Therefore he must use his imagination but he must nevertheless carefully control it by the exercise of *good* judgment. Second, only if the "guess" is correct and leads to solution, or at least puts him on the road to solution, is it a *good* guess. But, while realizing the usefulness and the time and labor-saving features of a solution by assuming a probable word, the cryptanalyst should exercise discretion in regard to how long he may continue in his efforts with this method. Sometimes he may actually waste time by adhering to the method too long, if straightforward, methodical analysis will yield results more quickly.

c. Obviously, the "probable-word" method has much more applicability when working upon material the general nature of which is known, than when working upon more or less isolated communications exchanged between correspondents concerning whom or whose activities

[7] This process does not involve anything more mysterious than ordinary, logical reasoning; there is nothing of the subnormal or supernormal about it. If cryptanalytic success seems to require processes akin to those of medieval magic, if "hocus-pocus" is much to the fore, the student should begin to look for items that the claimant of such success has carefully hidden from view, for the mystification of the uninitiated. (See Par. 33 in this connection.)

nothing is known. For in the latter case there is little or nothing that the imagination can seize upon as a background or basis for the assumptions.[8]

d. Very frequently, the choice of probable words is aided or limited by the number and positions of repeated letters. These repetitions may be *patent*—that is, externally visible in the cryptographic text as it originally stands—or they may be *latent*—that is, externally invisible but susceptible of being made patent as a result of the analysis. For example, in a monoalphabetic substitution cipher, such as that discussed in the preceding paragraph, the repeated letters are directly exhibited in the cryptogram; later the student will encounter many cases in which the repetitions are latent, but are made patent by the analytical process. When the repetitions are patent, then the *pattern* or *formula* to which the repeated letters conform is of direct use in assuming plain-text words; and when the text is in word-lengths, the pattern is obviously of even greater assistance. Suppose the cryptanalyst is dealing with military text, in which case he may expect such words as DIVISION, BATTALION, etc., to be present in the text. The positions of the repeated letter I in DIVISION, of the reversible digraph AT, TA in BATTALION, and so on, constitute for the experienced cryptanalyst tell-tale indications of the presence of these words, even when the text is not divided up into its original word lengths.

e. The important aid that a study of word patterns can afford in cryptanalysis warrants the use of definite terminology and the establishment of certain data having a bearing thereon. The phenomenon herein under discussion, namely, that many words are of such construction as regards the number and positions of repeated letters as to make them readily identifiable, will be termed *idiomorphism* (from the Greek "idios"=one's own, individual, peculiar+"morphe"=form). Words which show this phenomenon will be termed *idiomorphic*. It will be useful to deal with the idiomorphisms symbolically and systematically as described below.

f. When dealing with cryptograms in which the word lengths are determined or specifically shown, it is convenient to indicate their lengths and their repeated letters in some easily recognized manner or by formulas. This is exemplified, in the case of the word DIVISION, by the formula ABCBDBEF; in the case of the word BATTALION, by the formula ABCCBDEFG. If the cryptanalyst, during the course of his studies, makes note of striking formulas he has encountered, with the words which fit them, after some time he will have assembled a quite valuable body of data. And after more or less complete lists of such formulas have been established in some systematic arrangement, a rapid comparison of the idiomorphs in a specific cryptogram with those in his lists will be feasible and will often lead to the assumption of the correct word. Such lists can be arranged according to word length, as shown herewith:

```
3/aba  : DID, EVE, EYE.
  abb  : ADD, ALL, ILL, OFF, etc.
4/abac : ARAB, AWAY, etc.
  abca : AREA, BOMB, DEAD, etc.
  abbc : . . .
  abcb : . . .
  etc.   etc.
```

[8] General Givierge in his *Cours de Cryptographie* (p. 121) says: "However, expert cryptanalysts often employ such details as are cited above [in connection with assuming the presence of 'probable words'], and the experience of the years 1914 to 1918, to cite only those, prove that in practice one often has at his disposal elements of this nature, permitting assumptions much more audacious than those which served for the analysis of the last example. The reader would therefore be wrong in imagining that such fortuitous elements are encountered only in cryptographic works where the author deciphers a document that he himself enciphered. Cryptographic correspondence, if it is extensive, and if sufficiently numerous working data are at hand, often furnishes elements so complete that an author would not dare use all of them in solving a problem for fear of being accused of obvious exaggeration."

g. When dealing with cryptographic text in which the lengths of the words are not indicated or otherwise determinable, lists of the foregoing nature are not so useful as lists in which the words (or parts of words) are arranged according to the intervals between identical letters, in the following manner:

1 Interval	2 Intervals	3 Intervals	Repeated digraphs
–DiD–	AbbAcy	AbeyAnce	COCOa
–EvE–	ArAbiA	hAbitAble	dERER
–EyE–	AbiAtive	lAborAtory	ICICle
dIvIsion	AboArd	AbreAst	ININg
revIsIon	–AciA–	AbroAd	bAGgAGe
etc.	etc.	etc.	etc.

34. Solution of additional cryptograms produced by the same cipher component.—*a.* To return, after a rather long digression, to the cryptogram solved in pars. 28–31, once the cipher component of a cipher alphabet has been reconstructed, subsequent messages which have been enciphered by means of the same cipher component may be solved very readily, and without recourse to the principles of frequency, or application of the probable-word method. It has been seen that the illustrative cryptogram treated in paragraphs 24–31 was enciphered by juxtaposing the cipher component against the normal sequence so that $A_p=S_c$. It is obvious that the cipher component may be set against the plain component at any one of 26 different points of coincidence, each yielding a different cipher alphabet. After a cipher component has been reconstructed, however, it becomes a *known* sequence, and the method of converting the cipher letters into their plain-component equivalents and then completing the plain-component sequence begun by each equivalent can be applied to solve any cryptogram which has been enciphered by that cipher component.

b. An example will serve to make the process clear. Suppose the following message, passing between the same two stations as before, was intercepted shortly after the first message had been solved:

I Y E W K C E R N W O F O S E L F O O H E A Z X X

It is assumed that the same cipher component was used, but with a different key letter. First the initial two groups are converted into their plain-component equivalents by setting the cipher component against the normal sequence at any arbitrary point of coincidence. The initial letter of the former may as well be set against A of the latter, with the following result:

Plain..... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher..... L E A V N W O R T H B C D F G I J K M P Q S U X Y Z

Cryptogram.... I Y E W K C E R N W . . .
Equivalents.... P Y B F R L B H E F . . .

The normal sequence initiated by each of these conversion equivalents is now completed, with the results shown in Fig. 15. Note the plain-text generatrix, CLOSEYOURS, which manifests itself without further analysis. The rest of the message may be read either by continuing the

same process, or, what is even more simple, the key letter of the message may now be determined quite readily and the message deciphered by its means.

I	Y	E	W	K	C	E	R	N	W
P	Y	B	F	R	L	B	H	E	F
Q	Z	C	G	S	M	C	I	F	G
R	A	D	H	T	N	D	J	G	H
S	B	E	I	U	O	E	K	H	I
T	C	F	J	V	P	F	L	I	J
U	D	G	K	W	Q	G	M	J	K
V	E	H	L	X	R	H	N	K	L
W	F	I	M	Y	S	I	O	L	M
X	G	J	N	Z	T	J	P	M	N
Y	H	K	O	A	U	K	Q	N	O
Z	I	L	P	B	V	L	R	O	P
A	J	M	Q	C	W	M	S	P	Q
B	K	N	R	D	X	N	T	Q	R
*C	L	O	S	E	Y	O	U	R	S
D	M	P	T	F	Z	P	V	S	T
E	N	Q	U	G	A	Q	W	T	U
F	O	R	V	H	B	R	X	U	V
G	P	S	W	I	C	S	Y	V	W
H	Q	T	X	J	D	T	Z	W	X
I	R	U	Y	K	E	U	A	X	Y
J	S	V	Z	L	F	V	B	Y	Z
K	T	W	A	M	G	W	C	Z	A
L	U	X	B	N	H	X	D	A	B
M	V	Y	C	O	I	Y	E	B	C
N	W	Z	D	P	J	Z	F	C	D
O	X	A	E	Q	K	A	G	D	E

c. In order that the student may understand without question just what is involved in the latter step, that is, discovering the key letter after the first two or three groups have been deciphered by the conversion-completion process, the foregoing example will be used. It was noted that the first cipher group was finally deciphered as follows:

Cipher..... I Y E W K
Plain..... C L O S E

Now set the cipher component against the normal sequence so that $C_p = I_c$. Thus:

Plain..... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher..... F G I J K M P Q S U X Y Z L E A V N W O R T H B C D

It is seen here that when $C_p = I_c$ then $A_p = F_c$. This is the key for the entire message. The decipherment may be completed by direct reference to the foregoing cipher alphabet. Thus:

Cipher..... I Y E W K C E R N W O F O S E L F O O H E A Z X X
Plain..... C L O S E Y O U R S T A T I O N A T T W O P M X X

Message: CLOSE YOUR STATION AT TWO PM

d. The student should make sure that he understands the fundamental principles involved in this quick solution, for they are among the most important principles in cryptanalytics. How useful they are will become clear as he progresses into more and more complex cryptanalytic studies.

SECTION VII

MULTILITERAL SUBSTITUTION WITH SINGLE-EQUIVALENT CIPHER ALPHABETS

	Paragraph
Analysis of multiliteral, monoalphabetic substitution systems	35
Historically interesting examples	36

35. Analysis of multiliteral, monoalphabetic substitution systems.—*a.* Substitution methods in general may be classified into uniliteral and multiliteral systems.[1] In the former there is a strict "one-to-one" correspondence between the length of the units of the plain and those of the cipher text; that is, each letter of the plain text is replaced by a single character in the cipher text. In the latter this correspondence is no longer $1_p:1_c$ but may be $1_p:2_c$, where each letter of the plain text is replaced by a combination of two characters in the cipher text; or $1_p:3_c$, where a 3-character combination in the cipher text represents a single letter of the plain text, and so on. A cipher in which the correspondence of the $1_p:1_c$ type is termed uniliteral in character; one in which it is of the $1_p:2_c$ type, biliteral; $1_p:3_c$, triliteral, and so on. Those beyond the $1_p:1_c$ type are classed together as *multiliteral*.

b. When a multiliteral system employs biliteral equivalents, the cipher alphabet is said to be bipartite. Such alphabets are composed of a set of 25 or 26 combinations of a limited number of characters taken in pairs. An example of such an alphabet is the following.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M
Cipher	WW	WH	WI	WT	WE	HW	HH	HI	HT	HT	HE	IW	IH
Plain	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	II	IT	IE	TW	TH	TI	TT	TE	EW	EH	EI	ET	EE

This alphabet is derived from the square shown in Fig. 15.

(2)

(1)		W	H	I	T	E
	W	A	B	C	D	E
	H	F	G	H	I–J	K
	I	L	M	N	O	P
	T	Q	R	S	T	U
	E	V	W	X	Y	Z

FIGURE 15.

c. If a message is enciphered by means of the foregoing bipartite alphabet the cryptogram is still monoalphabetic in character. A frequency distribution based upon pairs of letters will

[1] See Sec. VII, *Advanced Military Cryptography*.

obviously have all the characteristics of a simple, uniliteral distribution for a monoalphabetic substitution cipher.

d. Ciphers of this type, as well as of those of the multiliteral (triliteral, quadraliteral, . . .) type are readily detected externally by virtue of the fact that the cryptographic text is composed of but a very limited number of different characters. They are handled in exactly the same manner as are uniliteral, monoalphabetic substitution ciphers. So long as the same character, or combination of characters, is always used to represent the same plain-text letter, and so long as a given letter of the plain text is always represented by the same character or combination of characters, the substitution is strictly monoalphabetic and can be handled in the simple manner described under Par. 31 of this text.

e. An interesting example in which the cipher equivalents are quinqueliteral groups and yet the resulting cipher is strictly monoalphabetic in character is found in the cipher system invented by Sir Francis Bacon over 300 years ago. Despite its antiquity the system possesses certain features of merit which are well worth noting. Bacon[2] proposed the following cipher alphabet, composed of permutations of two elements taken five at a time:[3]

A=aaaaa	I–J=abaaa	R=baaaa
B=aaaab	K=abaab	S=baaab
C=aaaba	L=ababa	T=baaba
D=aaabb	M=ababb	U–V=baabb
E=aabaa	N=abbaa	W=babaa
F=aabab	O=abbab	X=babab
G=aabba	P=abbba	Y=babba
H=aabbb	Q=abbbb	Z=babbb

If this were all there were to Bacon's invention it would be hardly worth bringing to attention. But what he pointed out, with great clarity and simple examples, was how such an alphabet might be used to convey a secret message by enfolding it in an innocent, external message which might easily evade the strictest kind of censorship. As a very crude example, suppose that a message is written in capital and lower case letters, any capital letter standing for an "a" element of the cipher alphabet, and any small letter, for a "b" element. Then the external sentence "All is well with me today" can be made to contain the secret message "Help." Thus:

A L l	i s	W E l L	W I t H	m E	T o d a Y
a a b	b b	a a b a	a	a b a	b a	a b b b a
H		E		L		P

Instead of employing such an obvious device as capital and small letters, suppose that an "a" element be indicated by a very slight shading, or a very slightly heavier stroke. Then a secret message might easily be thus enfolded within an external message of exactly opposite meaning. The number of possible variations of this basic scheme is very high. The fact that the characters

[2] For a true picture of this cipher, the explanation of which is often distorted beyond recognition even by cryptographers, see Bacon's own description of it as contained in his *De Augmentis Scientiarum* (*The Advancement of Learning*), as translated by any first-class editor, such as Gilbert Watts (1640) or Ellis, Spedding, and Heath (1857, 1870). The student is cautioned, however, not to accept as true any alleged "decipherments" obtained by the application of Bacon's cipher to literary works of the 16th century. These readings are purely subjective.

[3] In the 16th Century, the letters I and J were used interchangeably, as were also U and V. Bacon's alphabet was called by him a "biliteral alphabet" because it employs permutations of two letters. But from the cryptanalytic standpoint the significant point is that each plain-text letter is represented by a 5-character equivalent. Hence, present terminology requires that this alphabet be referred to as a *quinqueliteral alphabet.*

of the cryptographic text are hidden in some manner or other has, however, no effect upon the strict monoalphabeticity of the scheme.

36. Historically interesting examples.—*a.* Two examples of historical interest will be cited in this connection as illustrations. During the campaign for the presidential election of 1876 many cipher messages were exchanged between the Tilden managers and their agents in several states where the voting was hotly contested. Two years later the New York Tribune [4] exposed many irregularities in the campaign by publishing the decipherments of many of these messages. These decipherments were achieved by two investigators employed by the Tribune, and the plain text of the messages seems to show that illegal attempts and measures to carry the election for Tilden were made by his managers. Here is one of the messages:

JACKSONVILLE, Nov. 16 (1876).

GEO. F. RANEY, Tallahassee.

P p y y e m n s n y y y p i m a s h n s y y s s i t e p a a e n s h n s
p e n s s h n s m m p i y y s n p p y e a a p i e i s s y e s h a i n s s s p
e e i y y s h n y n s s s y e p i a a n y i t n s s h y y s p y y p i n s y y
s s i t e m e i p i m m e i s s e i y y e i s s i t e i e p y y p e e i a a s s
i m a a y e s p n s y y i a n s s s e i s s m m p p n s p i n s s n p i n s i m
i m y y i t e m y y s s p e y y m m n s y y s s i t s p y y p e e p p p m a
a a y y p i i t

L'Engle goes up tomorrow.

DANIEL.

Examination of the message discloses that only ten different letters are used. It is probable, therefore, that what one has here is a cipher which employs a bipartite alphabet and in which combinations of two letters represent single letters of the plain text. The message is therefore rewritten in pairs and substitution of arbitrary letters for the pairs is made, as seen below:

PP	YY	EM	NS	NY	YY	PI	MA	SH	NS	YY	SS	etc.
A	B	C	D	E	B	F	G	H	D	B	I	etc.

A triliteral frequency distribution is then made and analysis of the message along the lines illustrated in the preceding section of this text yields solution, as follows:

JACKSONVILLE, *Nov. 16.*

GEO. F. RANEY, *Tallahassee:*

Have Marble and Coyle telegraph for influential men from Delaware and Virginia. Indications of weakening here. Press advantage and watch Board. L'Engle goes up tomorrow.

DANIEL.

b. The other example, using numbers, is as follows:

JACKSONVILLE, Nov. 17.

S. PASCO and E. M. L'ENGLE:

84	55	84	25	93	34	82	31	31	75	93	82	77	33	55	42
93	20	93	66	77	66	33	84	66	31	31	93	20	82	33	66
52	48	44	55	42	82	48	89	42	93	31	82	66	75	31	93

DANIEL.

[4] New York Tribune, Extra No. 44, *The Cipher Dispatches*, New York, 1879.

There were, of course, several messages of like nature, and examination disclosed that only 26 different numbers in all were used. Solution of these ciphers followed very easily, the decipherment of the one given above being as follows:

JACKSONVILLE, *Nov. 17.*

S. PASCO and E. M. L'ENGLE:

Cocke will be ignored, Eagan called in. Authority reliable.

DANIEL.

c. The Tribune experts gave the following alphabets as the result of their decipherments:

AA=0	EN=Y	IT=D	NS=E	PP=H	SS=N
AI=U	EP=C	MA=B	NY=M	SH=L	YE=F
EI=I	IA=K	MM=G	PE=T	SN=P	YI=X
EM=V	IM=S	NN=J	PI=R	SP=W	YY=A
20=D	33=N	44=H	62=X	77=G	89=Y
25=K	34=W	48=T	66=A	82=I	93=E
27=S	39=P	52=U	68=F	84=C	96=M
31=L	42=R	55=0	75=B	87=V	99=J

They did not attempt to correlate these alphabets, or at least they say nothing about a possible relationship. The present author has, however, reconstructed the rectangle upon which these alphabets are based, and it is given below (fig. 16).

1st Letter or Number	2d Letter or Number									
	H 1	I 2	S 3	P 4	A 5	Y 6	M 7	E 8	N 9	T 0
H 1										
I 2					K		S			D
S 3	L		N	W					P	
P 4		R		H				T		
A 5		U			0					
Y 6		X				A		F		
M 7					B		G			
E 8		I		C			V		Y	
N 9			E			M			J	
T 0										

FIGURE 16.

It is amusing to note that the conspirators selected as their key a phrase quite in keeping with their attempted illegalities—HIS PAYMENT—for bribery seems to have played a considerable part in that campaign. The blank squares in the diagram probably contained proper names, numbers, etc.

SECTION VIII

MULTILITERAL SUBSTITUTION WITH MULTIPLE-EQUIVALENT CIPHER ALPHABETS

	Paragraph
Purpose of providing multiple-equivalent cipher alphabets	37
Solution of a simple example	38
Solution of more complicated example	39
A subterfuge to prevent decomposition of cipher text into component units	40

37. Purpose of providing multiple-equivalent cipher alphabets.—*a.* It has been seen that the characteristic frequencies of letters composing normal plain text, the associations they form in combining to form words, and the peculiarities certain of them manifest in such text all afford direct clues by means of which ordinary monoalphabetic substitution encipherments of such plain text may be more or less speedily solved. This has led to the introduction of simple methods for disguising or suppressing the manifestations of monoalphabeticity, so far as possible. Basically these methods are multiliteral and they will now be presented.

b. Multiliteral substitution may be of two types: (1) That wherein each letter of the plain text is represented by one and only one multiliteral equivalent. For example, in the Francis Bacon cipher described in Par. 35*e*, the letter K_p is invariably represented by the permutation abaab. For this reason this type of system may be more completely described as *monoalphabetic, multiliteral substitution with single-equivalent cipher alphabets.*

(2) That wherein, because of the large number of equivalents made available by the combinations and permutations of a limited number of elements, each letter of the plain text may be represented by several multiliteral equivalents which may be selected at random. For example, if 3-letter combinations are employed there are available 26^3 or 17,576 equivalents for the 26 letters of the plain text; they may be assigned in equal numbers of different equivalents for the 26 letters, in which case each letter would be representable by 676 different 3-letter equivalents; or they may be assigned on some other basis, for example, proportionately to the relative frequencies of plain-text letters. For this reason this type of system may be more completely described as *monoalphabetic, multiliteral substitution with multiple-equivalent cipher alphabets.* Some authors term such a system "simple substitution with multiple equivalents"; others term it *monoalphabetic substitution with variants.* For the sake of brevity, the latter designation will be employed in this text.

c. The primary object of monoalphabetic substitution with variants is, as has been mentioned above, to provide several values which may be employed at random in a simple substitution of cipher equivalents for the plain-text letters. In this connection, reference is made to Section X of *Elementary Military Cryptography,* wherein several of the most common methods for producing and using variants are set forth.

d. A word or two concerning the underlying theory from the cryptanalytic point of view of monoalphabetic substitution with variants, may not be amiss. Whereas in simple or single-equivalent, monoalphabetic substitution it is seen that—

(1) The same letter of the plain text is invariably represented by but one and always the same character of the cryptogram, and

(2) The same character of the cryptogram invariably represents one and always the same letter of the plain text;

In multiliteral substitution with multiple equivalents (monoalphabetic substitution with variants) it is seen that—

(1) The same letter of the plain text may be represented by one or more different characters of the cryptogram, but

(2) The same character of the cryptogram nevertheless invariably represents one and always the same letter of the plain text.

38. Solution of a simple example.—*a.* The following cryptogram has been enciphered by a set of four alphabets similar to the following:

A	B	C	D	E	F	G	H	I-J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	01	02	03	04	05	06	07
35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	26	27	28	29	30	31	32	33	34
68	69	70	71	72	73	74	75	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
87	88	89	90	91	92	93	94	95	96	97	98	99	00	76	77	78	79	80	81	82	83	84	85	86

The keyword here is TRIP[1]. In enciphering a message the equivalents are to be selected at random from among the four variants for each letter. The steps in solving a message produced by such a scheme will now be scrutinized.

CRYPTOGRAM

68321 09022 48057 65111 88648 42036 45235 09144 05764 22684
00225 57003 97357 14074 82524 40768 51058 93074 92188 47264
09328 04255 06186 79882 85144 45886 32574 55136 56019 45722
76844 68350 45219 71649 90528 65106 11886 44044 89669 70553
18491 06985 48579 33684 50957 70612 09795 29148 56109 08546
62062 65509 32800 32568 97216 44282 34031 84989 68564 53789
12530 77401 68494 38544 11368 87616 56905 20710 58864 67472
22490 09136 62851 24551 35180 14230 50886 44084 06231 12876
05579 58980 29503 99713 32720 36433 82689 04516 52263 21175
06445 72255 68951 86957 76095 67215 53049 08567 9730

b. Assuming that the foregoing remarks had not been made and that the cryptogram has just been submitted for solution with no information concerning it, the first step is to make a preliminary study to determine whether the cryptogram involves cipher or code. The cryptogram appears in 5-figure groups, which may indicate either cipher or code. A few remarks will be made at this point with reference to the method of determining whether a cryptogram composed of figure groups is in code or cipher, using the foregoing example.

c. In the first place, if the cryptogram contains an even number of digits, as for example 494 in the foregoing message, this leaves open the possibility that it may be cipher, composed of 247 pairs of digits; were the number of digits an exact *odd* multiple of five, such as 125, 135, etc., the possibility that the cryptogram is in code of the 5-figure group type must be considered. Next, a preliminary study is made to see if there are many repetitions, and what their characteristics

[1] The letter corresponding to the lowest number in each line of the diagram showing the cipher alphabets is a key letter. Thus, in the 1st line 01=T; in the 2d line 26=R; etc.

are. If the cryptogram is code of the 5-figure group type, then such repetitions as appear should *generally* be in whole groups of five digits, and they should be visible in the text just as the message stands, unless the code message has undergone encipherment also. If the cryptogram is in cipher, then the repetitions should extend beyond the 5-digit groupings; if they conform to any definite groupings at all they should for the most part contain even numbers of digits since each letter is probably represented by a pair of digits. If no clues of the foregoing nature are present, doubts will be dissolved by making a detailed study of frequencies.

d. A simple 4-part frequency distribution is therefore decided upon. Shall the alphabet be assumed to be a 25- or a 26-character one? If the former, then the 2-digit pairs from 01 to 00 fall into exactly four groups each corresponding to an alphabet. Since this is the most common scheme of drawing up such alphabets, let it be assumed to be true of the present case. The following distributions result from the breaking up of the text into 2-digit pairs.

01—///	26—///	51—𝍸	76—𝍸 /
02—	27—	52—𝍸	77—/
03—////	28—/	53—///	78—
04—/	29—/	54—	79—/
05—𝍸	30—///	55—////	80—///
06—𝍸 /	31—	56—𝍸	81—
07—///	32—𝍸 /	57—𝍸 /	82—////
08—	33—/	58—//	83—/
09—////	34—/	59—	84—𝍸 /
10—////	35—//	60—	85—𝍸 /
11—𝍸	36—𝍸	61—	86—///
12—///	37—/	62—//	87—
13—/	38—	63—	88—////
14—/	39—/	64—𝍸 /	89—𝍸
15—/	40—///	65—	90—𝍸 /
16—///	41—	66—/	91—///
17—	42—////	67—//	92—/
18—𝍸 /	43—/	68—𝍸 //	93—/
19—	44—𝍸 /	69—//	94—/
20—/	45—𝍸 /	70—/	95—///
21—//	46—///	71—/	96—
22—𝍸	47—	72—////	97—𝍸 /
23—//	48—///	73—	98—/
24—	49—𝍸	74—////	99—
25—/	50—𝍸	75—/	00—//

e. If the student will bring to bear upon this problem the principles he learned in Section V of this text, he will soon realize that what he now has before him are four, simple, monoalphabetic frequency distributions similar to those involved in a monoalphabetic substitution cipher using standard cipher alphabets. The realization of this fact immediately provides the clue to the next step: "fitting each of the distributions to the normal." (See Par. 17*b*). This can be

done without difficulty in this case (remembering that a 25-letter alphabet is involved and assuming that I and J are the same letter) and the following alphabets result:

01—I–J	26—U	51—N	76—E
02—K	27—V	52—O	77—F
03—L	28—W	53—P	78—G
04—M	29—X	54—Q	79—H
05—N	30—Y	55—R	80—I–J
06—O	31—Z	56—S	81—K
07—P	32—A	57—T	82—L
08—Q	33—B	58—U	83—M
09—R	34—C	59—V	84—N
10—S	35—D	60—W	85—O
11—T	36—E	61—X	86—P
12—U	37—F	62—Y	87—Q
13—V	38—G	63—Z	88—R
14—W	39—H	64—A	89—S
15—X	40—I–J	65—B	90—T
16—Y	41—K	66—C	91—U
17—Z	42—L	67—D	92—V
18—A	43—M	68—E	93—W
19—B	44—N	69—F	94—X
20—C	45—O	70—G	95—Y
21—D	46—P	71—H	96—Z
22—E	47—Q	72—I–J	97—A
23—F	48—R	73—K	98—B
24—G	49—S	74—L	99—C
25—H	50—T	75—M	00—D

f. The keyword is seen to be JUNE and the first few groups of the cryptogram decipher as follows:

68	32	10	90	22	48	05	76	51	11	88	64	84	20	36	45	23
E	A	S	T	E	R	N	E	N	T	R	A	N	C	E	O	F

g. From the detailed procedure given above, the student should be able to draw his own conclusions as to the procedure to be followed in solving cryptograms produced by methods which are more or less simple variations of that just discussed. In this connection he is referred to Section X of *Elementary Military Cryptography*, wherein a few of these variations are mentioned.

h. Possibly the most important of the variations is that in which a rectangle such as that shown in Fig. 17 is employed.

	1	2	3	4	5	6	7	8	9	0
1, 4, 7	A	B	C	D	E	F	G	H	I	J
2, 5, 8	K	L	M	N	O	P	Q	R	S	T
3, 6, 9	U	V	W	X	Y	Z	–	,	:	;

FIGURE 17

In the solution of cases of this kind, repetitions would play their usual role, with the modifications noted below in Par. 39. Once an entering wedge has been forced, through the identification of one or more repeated words such as BATTALION, DIVISION, etc., the entire enciphering rectangle would soon be reconstructed. It may be added that the frequency distribution for the text of a single long message or several short ones enciphered by such a system would show characteristic phenomena, the most important of which are, first, that the distribution for a rectangle such as shown in Fig. 17 would practically follow the normal and, second, that the distribution for the 2d digit of pairs would show more marked crests and troughs than the distribution for the 1st digit. For example, the initial digits 1, 4, and 7 (for the numbers 10–19, 40–49, and 70–79, inclusive) would apply to the distribution for the letters A to J, inclusive; the initial digits 2, 5, and 8 would apply to the distribution for the letters K to T, inclusive. The total weighted frequency values for these two groups of letters are about equal. Therefore, the frequencies of the initial digits 1, 2, 4, 5, 7, and 8 would be approximately equal. But consider the final digit 5 in the numbers 15, 45, 75, 25, 55, and 85; its total frequency is composed of the frequency of E_p plus the frequency of O_p; whereas in the case of the final digit 6, its total frequency is composed of the frequency of F_p plus the frequency of Q_p. The two cases would show a marked difference in frequency. Of course, the letters may be inserted within the enciphering rectangle in a keyword-mixed or even in a random order; the numbers may be applied to the rectangle in a random order. But these variations, while increasing the difficulty in solution, by no means make the latter as great as may be thought by the novice.

39. Solution of a more complicated example.—*a.* As soon as a beginner in cryptography realizes the consequences of the fact that letters are used with greatly varying frequencies in normal plain text, a brilliant idea very speedily comes to him. Why not disguise the natural frequencies of letters by a system of substitution using many equivalents, and let the numbers of equivalents assigned to the various letters be more or less in direct proportion to the normal frequencies of the letters? Let E, for example, have 13 or more equivalents; T, 10; N, 9; etc., and thus (he thinks) the enemy cryptanalyst can have nothing in the way of tell-tale or characteristic frequencies to use as an entering wedge.

b. If the text available for study is small in amount and if the variant values are wholly independent of one another, the problem can become exceedingly difficult. But in practical military communications such methods are rarely encountered, *because the volume of text is usually great enough to permit of the establishment of equivalent values.* To illustrate what is meant, suppose a set of cryptograms produced by the monoalphabetic-variant method described above shows the following two sets of groupings in the text:

SET A	SET B
12–37–02–79–68–13–03–37–77	71–12–02–51–23–05–77
82–69–03–79–13–68–23–37–35	11–82–51–02–03–05–35
82–69–51–16–13–13–78–05–35	11–91–02–02–23–37–35
91–05–02–01–68–42–78–37–77	97–12–51–03–78–69–77

An examination of these groupings would lead to the following tentative conclusions with regard to probable equivalents:

12, 82, 91	01, 16, 79	03, 23, 78
05, 37, 69	13, 42, 68	35, and 77
02, and 51		

The establishment of these equivalencies would sooner or later lead to the finding of additional sets of equal values. The completeness with which this can be accomplished will determine

the ease or difficulty of solution. Of course, if many equivalencies can be established the problem can then be reduced practically to monoalphabetic terms and a speedy solution can be attained.

c. Theoretically, the determination of equivalencies may seem to be quite an easy matter, but practically it may be very difficult, because the cryptanalyst can never be *certain* that a combination showing what may appear to be a variant value is really such, and is not a different word. For example, take the groups—

17–82–31–82–14–63, and
27–82–40–82–14–63

Here one might suspect that 17 and 27 represent the same letter, 31 and 40 another letter. But it happens that one group represents the word MANAGE, the other DAMAGE.

d. When reversible combinations are used as variants, the problem is perhaps a bit more simple. For example, using the accompanying Fig. 18 for encipherment, two messages with the same initial words, REFERENCE YOUR, may be enciphered as follows:

	K,Z	Q,V	B,H	M,R	D,L
W,S	N	H	A	O	E
F,X	D	T	M	F	P
G,J	Q	B	U	I	V
C,N	G	X	R	C	S
P,T	Z	L	Y	W	K

FIGURE 18.

	R	E	F	E	R	E	N	C	E	Y	O	U	R
(1)	N H W D R		X L S H C		D W W Z N		R S L H P			S R B J C		H	
(2)	C H D W R		X S L H N		D W Z W N		R L S H P			R W J B N		H	

The experienced cryptanalyst, noting the appearance of the very first few groups, assumes that he is here confronted with a case involving biliteral reversible equivalents, with variants.

e. The probable-word method of solution may be used, but with a slight variation introduced by virtue of the fact that, regardless of the system, *letters of low frequency in plain text remain infrequent.* Hence, suppose a word containing low-frequency letters, but in itself a rather common word strikingly idiomorphic in character is sought as a "probable word"; for example, words such as CAVALRY, ATTACK, and PREPARE. Writing such a word on a slip of paper, it is slid one interval at a time under the text, which has been marked so that the high and low-frequency characters are indicated. Each coincidence of a low-frequency letter of the text with a low-frequency letter of the assumed word is examined carefully to see whether the adjacent text letters correspond in frequency with the other letters of the assumed word; or, if the latter presents repetitions, whether there are correspondences between repetitions in the text and those in the word. Many trials are necessary but this method will produce results when the difficulties are otherwise too much for the cryptanalyst to overcome.

40. A subterfuge to prevent decomposition of cipher text into component units.—*a.* A few words should be added with regard to certain subterfuges which are sometimes encountered in monoalphabetic substitution with variants, and which, if not recognized in time, cause considerable delays. These have to deal with the insertion of nulls so as to prevent the cryptanalyst from breaking up the text into its real cryptographic units. The student should take careful

note of the last phrase; the mere insertion of symbols having the same characteristics as the symbols of the cryptographic text, except that they have no meaning, is not what is meant. This class of nulls rarely achieves the purpose for which they are intended. What is really meant can best be explained in connection with an example. Suppose that a 5 x 5 checkerboard design with the row and column indicators shown in Fig. 19 is adopted for encipherment. Normally, the cipher units would consist of 2-letter combinations of the indicators, invariably giving the row indicator first (by agreement).

	V	G	I	W	D
	A	H	P	S	M
	T	O	E	B	N
	F	U	R	L	C
V,A,T,F	A	B	C	D	E
G,H,O,U	F	G	H	I–J	K
I,P,E,R	L	M	N	O	P
W,S,B,L	Q	R	S	T	U
D,M,N,C	V	W	X	Y	Z

FIGURE 19.

The phrase COMMANDER OF SPECIAL TROOPS might be enciphered thus:

C	O	M	M	A	N	D	E	R	O	F	. . .
VI	EB	PH	IU	FT	IE	AB	TM	WO	PW	GT	. . .

These would normally then be arranged in 5-letter groups, thus:

V I E B P H I U F T I E A B T M W O P W G T . . .

b. It will be noted, however, that only 20 of the 26 letters of the alphabet have been employed as row and column indicators, leaving J, K, Q, X, Y, and Z unused. Now, suppose these five letters are used as nulls, *not in pairs, but as individual letters inserted at random* just before the real text is arranged in 5-letter groups. Occasionally, a pair of nulls is inserted. Thus, for example:

V I E X B P H K I U F J X T I E A J B T M W O Q P W G K T Y

The cryptanalyst, after some study, suspecting a biliteral cipher, proceeds to break up the text into pairs:

VI EX BP HK IU FJ XT IE AJ BT MW OQ PW GK TY

Compare this set of 2-letter combinations with the correct set. Only 4 of the 15 pairs are "proper" units. It is easy to see that without a knowledge of the *existence* of the nulls, and even with a knowledge, if he does not know *which* letters are nulls, the cryptanalyst would be confronted with a problem for the solution of which a fairly large amount of text might be necessary. The careful employment of the variants also very materially adds to the security of the method because repetitions can be rather effectively suppressed.

c. From the cryptographic standpoint, the fact that in this system the cryptographic text is more than twice as long as the plain text constitutes a serious disadvantage. From the cryptanalytic standpoint, the masking of the cipher units constitutes the most important source of strength of the system; this, coupled with the use of variants, makes it a bit more difficult system to solve, despite its monoalphabeticity.

SECTION IX

POLYGRAPHIC SUBSTITUTION SYSTEMS

	Paragraph
Monographic and polygraphic substitution systems	41
Tests for identifying digraphic substitution	42
General procedure in the analysis of digraphic substitution ciphers	43
Analysis of digraphic substitution ciphers based upon 4-square checkerboard designs	44
Analysis of ciphers based upon other types of checkerboard designs	45
Analysis of the Playfair cipher system	46

41. Monographic and polygraphic substitution systems.—*a.* The student is now referred to Sections VII and VIII of *Advanced Military Cryptography*, wherein polygraphic systems of substitution are discussed from the cryptographic point of view. These will now be discussed from the cryptanalytic point of view.

b. Although the essential differences between polyliteral and polygraphic substitution are treated with some detail in Section VII of *Advanced Military Cryptography*, a few additional words on the subject may not be amiss at this point.

c. The two primary divisions of substitution systems into (1) uniliteral and multiliteral methods and into (2) monographic and polygraphic methods are both based upon considerations as to the *number of elements* constituting the plain-text and the equivalent cipher-text units. In uniliteral as well as in monographic substitution, each plain-text unit consists of a single element and each cipher-text unit consists of a single element. The two terms uniliteral and monographic are therefore identical in significance, as defined cryptographically. It is when the terms multiliteral and polygraphic are examined that an essential difference is seen. In multiliteral substitution the plain-text unit always consists of a single element (one letter) and the cipher-text unit consists of a group of two or more elements; when biliteral, it is a pair of elements, when triliteral, it is a set of three elements, and so on. In what will herein be designated as true or complete polygraphic substitution the plain-text unit consists of two or more elements forming an *indivisible compound;* the cipher-text unit usually consists of a corresponding number of elements.[1] When the number of elements comprising the plain-text units is fixed and always two, the system is *digraphic;* when it is three, the system is *trigraphic;* when it is four, *tetragraphic;* and so on.[2] It is important to note that in true or complete polygraphic substitution the elements combine to form indivisible compounds having properties different from those of either of the constituent letters. For example, in uniliteral substitution AB_p may yield XY_c and AC_p may yield XZ_c; but in true digraphic substitution $\overline{AB}_p$ may yield $\overline{XY}_c$ and $\overline{AC}_p$ may yield $\overline{QN}_c$. A difference in identify of one letter affects the whole result.[3] An analogy is found in chemistry, when two elements combine to form a molecule, the latter usually having properties quite different from those of either of the constituent elements. For example: sodium, a metal, and

[1] The qualifying adverb "usually" is employed because this correspondence is not essential. For example, if one should draw up a set of 676 arbitrary single signs, it would be possible to represent the 2-letter pairs from AA to ZZ by single symbols. This would still be a digraphic system.

[2] In this sense a code system is merely a polygraphic substitution system in which the number of elements constituting the plain-text units is variable.

[3] For this reason the two letters are marked by a ligature, that is, by a bar across their tops.

chlorine, a gas, combine to form sodium chloride, common table salt. Furthermore, sodium and fluorine, also a gas similar in many respects to chlorine, combine to form sodium fluoride, which is much different from table salt. Partial and pseudo-polygraphic substitution will be treated under subparagraphs *d* and *e* below.

d. Another way of looking at polygraphic substitution is to regard the elements comprising the plain-text units as being enciphered individually and polyalphabetically by a fairly large number of separate alphabets. For example, in a digraphic system in which 676 pairs of plain-text letters are representable by 676 cipher-text pairs assigned at random, this is equivalent to having a set of 26 different alphabets for enciphering one member of the pairs, and another set of 26 different alphabets for enciphering the other member of the pairs. According to this viewpoint the different alphabets are brought into play by the particular combination of letters forming each plain-text pair. This is, of course, quite different from systems wherein the various alphabets are brought into play by more definite rules; it is perhaps this very absence of definite rules guiding the selection of alphabets which constitutes the cryptographic strength of this type of polygraphic system.

e. When regarded in the light of the preceding remarks, certain systems which at first glance seem to be polygraphic, in that groupings of plain-text letters are treated as units, on closer inspection are seen to be only partially polygraphic, or pseudo-polygraphic in character. For example, in a system in which encipherment is by pairs and yet one of the letters in each pair is enciphered monoalphabetically, the other letter, polyalphabetically, the method is only *pseudo*-polygraphic. Cases of this type are shown in Section VII of *Advanced Military Cryptography.* Again, in a system in which encipherment is by pairs and the encipherments of the left-hand and right-hand members of the pairs show group relationships, this is not pseudo-polygraphic but only *partially* polygraphic. Cases of this type are also shown in the text referred to above.

f. The fundamental purpose of polygraphic substitution is again the suppression of the frequency characteristics of plain text, just as is the case in monoalphabetic substitution with variants; but here this is accomplished by a different method, the latter arising from a somewhat different approach to the problem involved in producing cryptographic security. When the substitution involves replacement of *single* letters in a monoalphabetic system, the cryptogram can be solved rather readily. Basically the reason for this is that the principles of frequency and the laws of probability, applied to individual units of the text (single letters), have a very good opportunity to manifest themselves. A given volume of text of say n plain-text letters, enciphered purely monoalphabetically, affords n cipher characters, and the same number of cipher units. The same volume of text, enciphered digraphically, still affords n cipher characters but only $\frac{n}{2}$ cipher units. Statistically speaking, the sample within which the laws of probability now apply has been cut in half. Furthermore, from the point of view of frequency, the very noticeable diversity in the frequencies of individual letters, leading to the marked crests and troughs of the uniliteral frequency distribution, is no longer so strikingly in evidence in the frequencies of digraphs. Therefore, although true digraphic encipherment, for example, cuts the cryptographic textual units in half, the difficulty of solution is not doubled, but, if a matter of judgment arising from practical experience can be expressed or approximated mathematically, squared or cubed.

g. Sections VII and VIII of *Advanced Military Cryptography* show various methods for the derivation of polygraphic equivalents and for handling these equivalents in cryptographing and decryptographing messages. The most practicable of those methods are digraphic in character and for this reason their solution will be treated in a somewhat more detailed manner than will trigraphic methods. The latter can be passed over with the simple statement that their analysis requires much text to permit of solution by the frequency method, and hard labor. Fortunately, they are infrequently encountered because they are difficult to manipulate without extensive

tables.[4] If the latter are required they must be compiled in the form of a book or pamphlet. If one is willing to go that far, one might as well include in such document more or less extensive lists of words and phrases, in which case the system falls under the category of code and not cipher.

42. Tests for identifying digraphic substitution.—*a.* The tests which are applied to determine whether a given cryptogram is digraphic in character are usually rather simple. If there are many repetitions in the cryptogram and yet the uniliteral-frequency distribution gives no clear-cut indications of monoalphabeticity; if most of the repetitions contain an even number of letters; and if the cryptogram contains an even number of letters, it may be assumed to be digraphic in nature.

b. The student should first try to determine whether the substitution is completely digraphic, or only partially digraphic, or pseudo-digraphic in character. As mentioned above, there are cases in which, although the substitution is effected by taking pairs of letters, one of the members of the pairs is enciphered monoalphabetically, the other member, polyalphabetically. A distribution based upon the letters in the odd positions and one based upon those in the even positions should be made. If one of these is clearly monoalphabetic, then this is evidence that the message represents a case of pseudo-digraphism of the type here described. By attacking the monoalphabetic portion of the messages, solution can soon be reached by slight variation of the usual method, the polyalphabetic portion being solved by the aid of the context and considerations based upon the probable nature of the substitution chart. (See Tables 2, 3, and 4 of *Advanced Military Cryptography*.) It will be noted that the charts referred to show definite symmetry in their construction.

c. On the other hand, if the foregoing steps prove fruitless, it may be assumed that the cryptogram is completely digraphic in character.

d. Just as certain statistical tests may be applied to a cryptogram to establish its monoalphabeticity, so also may a statistical test be applied to a cryptogram for the purpose of establishing its digraphicity. The nature of this test and its method of application will be discussed in a subsequent text.

43. General procedure in the analysis of digraphic substitution ciphers.—*a.* The analysis of cryptograms which have been produced by digraphic substitution is accomplished largely by the application of the simple principles of frequency of digraphs, with the additional aid of such special circumstances as may be known to or suspected by the cryptanalyst. The latter refer to peculiarities which may be the result of the particular method employed in obtaining the equivalents of the plain-text digraphs in the cryptographing process. In general, however, only if there is sufficient text to disclose the normal phenomena of repetition will solution be feasible or possible.

b. However, when a digraphic system is employed in regular service, there is little doubt but that traffic will rapidly accumulate to an amount more than sufficient to permit of solution by simple principles of frequency. Sometimes only two or three long messages, or a half dozen of average length are sufficient. For with the identification of only a few cipher digraphs, larger portions of messages may be read because the skeletons of words formed from the few high-frequency digraphs very definitely limit the values that can be inserted for the intervening unidentified digraphs. For example, suppose that the plain-text digraphs TH, ER, IN, IS, OF, NT, and TO have been identified by frequency considerations, corroborated by a tentatively identified long repetition; and suppose also that the enemy is known to be using a quadricular

[4] A patent has been granted upon a rather ingenious machine for automatically accomplishing true polygraphic substitution, but it has not been placed upon the market. See U. S. Patent No. 1845947 issued in 1932 to Weisner and Hill. In U. S. Patent No. 1515680 issued to Henkels in 1924, there is described a mechanism which also produces polygraphic substitution.

table of 676 cells containing digraphs showing reciprocal equivalence between plain and cipher-text digraphs. Suppose the message begins as follows (in which the assumed values have been inserted):

XQ	VO	ZI	LK	AP	OL	ZX	PV	QN	IK	OL	UK	AL	HN	LK	VL
FO		TH	IN		NT		RE			NT	NO			IN	

BN	OZ	KU	DY	EL	LE	YW
SI		ON	TO			

The words FOURTH INFANTRY REGIMENT are readily recognized. The reciprocal pairs EL_c and LE_c suggest ATTACK. The beginning of the message is now completely disclosed: FOURTH INFANTRY REGIMENT NOT YET IN POSITION TO ATTACK. The values more or less automatically determined are $VO_c=UR_p$, $AL_c=TY_p$, $HN_c=ET_p$, $VL_c=PO_p$, $OZ_c=TI_p$, $YW_c=CK_p$.

c. Once a good start has been made and a few words have been solved, subsequent work is quite simple and straightforward. A knowledge of enemy correspondence, including data regarding its most common words and phrases, is of great assistance in breaking down new digraphic tables of the same nature but with different equivalents.

d. The foregoing remarks also apply to the details of solution in cases of partially digraphic substitution.

44. Analysis of digraphic substitution ciphers based upon 4-square checkerboard designs.— *a.* In Section VIII of *Advanced Military Cryptography* there are shown various examples of digraphic substitution based upon the use of checkerboard designs. These may be considered cases of partially digraphic substitution, in that in the checkerboard system there are certain relationships between plain-text digraphs having common elements and their corresponding cipher-text digraphs, which will also have common elements. For example, take the following 4-square checkerboard design:

	B	W	G	R	M	O	P	A	U	L	
	N	Y	V	X	E	H	Z	Q	D	F	
1	S	I	C	T	K	K	I	T	S	C	3
	U	P	L	A	O	M	W	R	B	G	
	D	Z	F	Q	H	E	Y	X	N	V	
	W	A	L	E	S	C	X	K	P	B	
	F	H	U	I	T	O	M	Y	D	V	
2	P	X	B	K	C	S	A	E	W	L	4
	N	Z	R	Q	G	G	Z	Q	N	R	
	D	M	V	Y	O	T	H	I	F	U	

FIGURE 20.

Here $BC_p=OW_c$, $BO_p=OF_c$, $BS_p=OP_c$, $BG_p=ON_c$ and $BT_p=OD_c$. In each case when B_p is the initial letter of the plain-text pair, the initial letter of the cipher-text equivalent is O_c. This, of course, is the direct result of the method; it means that the encipherment is monoalphabetic for the

first half of each of these *five* plain-text pairs, polyalphabetic for the second half. This relationship holds true for *four* other groups of pairs beginning with B_p. In other words, there are five alphabets employed, not 25. Thus, this case differs from the case discussed under Par. 42*b* only in that the monoalphabeticity is not complete for one-half of all the pairs, but only among the members of certain groups of pairs. In a completely digraphic system using a 676-cell randomized square, such relationships are entirely absent and for this reason the system is cryptographically more secure than the checkerboard system.

b. From the foregoing, it is clear that when solution has progressed sufficiently to disclose a few values, the insertion of letters within the cells of the checkerboard design to give the plain-text and cipher relationships indicated by the solved values immediately leads to the disclosure of additional values. Thus, the solution of only a few values soons leads to the breakdown of the entire checkerboard design.

c. (1) The following example will serve to illustrate the procedure. Let the message be as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
A.	H	F	C	A	P	G	O	Q	I	L	B	S	P	K	M	N	D	U	K	E	O	H	Q	N	F	B	O	R	U	N
B.	Q	C	L	C	H	Q	B	Q	B	F	H	M	A	F	X	S	I	O	K	O	Q	Y	F	N	S	X	M	C	G	Y
C.	X	I	F	B	E	X	A	F	D	X	L	P	M	X	H	H	R	G	K	G	Q	K	Q	M	L	F	E	Q	Q	I
D.	G	O	I	H	M	U	E	O	R	D	C	L	T	U	F	E	Q	Q	C	G	Q	N	H	F	X	I	F	B	E	X
E.	F	L	B	U	Q	F	C	H	Q	O	Q	M	A	F	T	X	S	Y	C	B	E	P	F	N	B	S	P	K	N	U
F.	Q	I	T	X	E	U	Q	M	L	F	E	Q	Q	I	G	O	I	E	U	E	H	P	I	A	N	Y	T	F	L	B
G.	F	E	E	P	I	D	H	P	C	G	N	Q	I	H	B	F	H	M	H	F	X	C	K	U	P	D	G	Q	P	N
H.	C	B	C	Q	L	Q	P	N	F	N	P	N	I	T	O	R	T	E	N	C	O	B	C	N	T	F	H	H	A	Y
I.	Z	L	Q	C	I	A	A	I	Q	U	C	H	T	P	C	B	I	F	G	W	K	F	C	Q	S	L	Q	M	C	B
J.	O	Y	C	R	Q	Q	D	P	R	X	F	N	Q	M	L	F	I	D	G	C	C	G	I	O	G	O	I	H	H	F
K.	I	R	C	G	G	G	N	D	L	N	O	Z	T	F	G	E	E	R	R	P	I	F	H	O	T	F	H	H	A	Y
L.	Z	L	Q	C	I	A	A	I	Q	U	C	H	T	P																

(2) The cipher having been tested for standard alphabets (by the method of completing the normal components) and found to give negative results, a uniliteral-frequency distribution is made. It is as follows:

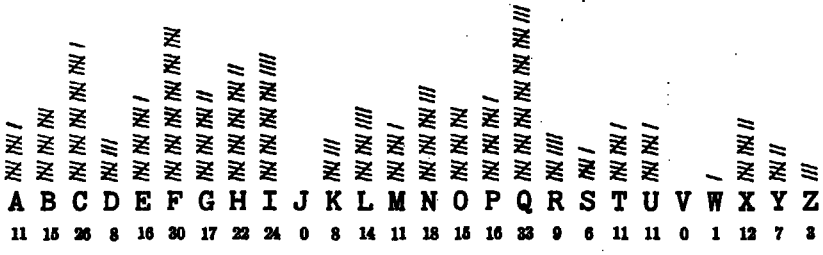


FIGURE 21.

(3) At first glance this may appear to the untrained eye to be a monoalphabetic frequency distribution but upon closer inspection it is noted that aside from the frequencies of four or five

letters the frequencies for the remaining letters are not very dissimilar. There are, in reality, no very marked crests and troughs, certainly not as many as would be expected in a monoalphabetic substitution cipher of equal length.

(4) The message having been carefully examined for repetitions of 4 or more letters, all of them are listed:

	Frequency	Located in lines
TFHHAYZLQCIAAIQUCHTP (20 letters)	2	H and K.
QMLFEQQIGOI (11 letters)	2	C and F.
XIFBEX (6 letters)	2	C and D.
FEQQ	3	C, D, F.
QMLF	3	C, F, J.
BFHM	2	B and G.
BSPK	2	A and E.
GOIH	2	D and J.

Since there are quite a few repetitions, two of considerable length, since all but one of them contain an even number of letters, and since the message also contains an even number of letters, 344, digraphic substitution is suspected. The cryptogram is transcribed in 2-letter groups, for greater convenience in study. It is as follows:

Message transcribed in pairs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B.	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C.	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM	LF	EQ	QI
D.	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E.	FL	BU	QF	CH	QO	QM	AF	TX	SY	CB	EP	FN	BS	PK	NU
F.	QI	TX	EU	QM	LF	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G.	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN
H.	CB	CQ	LQ	PN	FN	PN	IT	OR	TE	NC	CB	CN	TF	HH	AY
J.	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM	CB
K.	OY	CR	QQ	DP	RX	FN	QM	LF	ID	GC	CG	IO	GO	IH	HF
L.	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M.	ZL	QC	IA	AI	QU	CH	TP								

It is noted that all the repetitions listed above break up properly into digraphs except in one case, viz, FEQQ in lines C, D, and F. This seems rather strange, and at first thought one might suppose that a letter was dropped out or was added in the vicinity of the FEQQ in line D. But it is immediately seen that the FE QQ in line D has no relation at all to the .F EQ Q. in lines C and F, and that the F EQ Q in line D is merely an accidental repetition.

(5) A digraphic frequency distribution [5] is made and is shown in Fig. 22.

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A						3			2															2	
B						2										1		2		1					
C	1	5					4	3			1		1			2	1								
D															1					1			1		
E														1	2	2	1			1			2		
F		3			2						1		4												
G			1		1		1							3		1						1		1	
H						4		3				2		1	2	1									
I	3			2	1	2		3			1			2			1		1						
K					1	1	1							1						1					
L		1	1			3							1		1	1									
M			1										1							1			1		
N			1	1												1				1				1	
O								1								1	2							1	1
P				1			1			2			3												
Q			3			1			3	1		5	2	1		2				2				1	
R				1			1								1								1		
S											1												1	1	
T					1	4									2					1			2		
U					1								1												
V																									
W																									
X			1						2									1							
Y																									
Z											2														

FIGURE 22.

(6) The appearance of the foregoing distribution for this message is quite characteristic of that for a digraphic substitution cipher. There are many blank cells; although there are many cases in which a digraph appears only once, there are quite a few in which a digraph appears two or three times, four cases in which a digraph appears four times, and two cases in which a digraph appears five times. The absence of the letter J is also noted; this is often the case in a digraphic system based upon a checkerboard design.

[5] The distinction between "digraphic" and "biliteral" is based upon the following consideration. In a biliteral (or diliteral) distribution every two successive letters of the text would be grouped together to form a pair. For example, a biliteral distribution of ABCDEF would tabulate the pairs AB, BC, CD, DE, and EF. In a digraphic distribution only successive pairs of the text are tabulated. For example, ABCDEF would yield only AB, CD, and EF.

(7) In another common type of checkerboard system known as the Playfair cipher, described in Par. 46, one of the telltale indications besides the absence of the letter J is the absence of double letters, that is, two successive identical letters. The occurrence of the double letters GG, HH, and QQ in the message under investigation eliminates the possibility of its being a Playfair cipher. The simplest thing to assume is that a 4-square checkerboard is involved. One with normal alphabets in Sections 1 and 2 is therefore set down (Fig. 23*a*).

	A	B	C	D	E					
	F	G	H	I–J	K					
1	L	M	N	O	P					3
	Q	R	S	T	U					
	V	W	X	Y	Z					
						A	B	C	D	E
						F	G	H	I–J	K
4						L	M	N	O	P 2
						Q	R	S	T	U
						V	W	X	Y	Z

FIGURE 23*a*.

(8) The recurrence of the group QMLF, three times, and at intervals suggesting that it might be a sentence separator, leads to the assumption that it is the word STOP. The letters Q, M, L, and F are therefore inserted in the appropriate cells in Sections 3 and 4 of the diagram. Thus (Fig. 23*b*):

	A	B	C	D	E					
	F	G	H	I–J	K					
1	L	M	N	O	P					L 3
	Q	R	S	T	U				Q	
	V	W	X	Y	Z					
						A	B	C	D	E
						F	G	H	I–J	K
4				F		L	M	N	O	P 2
			M			Q	R	S	T	U
						V	W	X	Y	Z

FIGURE 23*b*.

148274—38——6

These placements seem logical. Moreover, in Section 3 the number of cells between L and Q is just one less than enough to contain all the letters M to P, inclusive, and suggests that either N or O is in the keyword portion of the sequence, that is, near the top of Section 3. Without making a commitment in the matter, suppose both N and O, for the present, be inserted in the cell between M and P. Thus (Fig. 23*c*):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
						F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*c*.

(9) Now, if the placement of P in Section 3 is correct, the cipher equivalent of TH_p will be $P\Theta_s$, and there should be a group of adequate frequency to correspond. Noting that PN_s occurs three times, it is assumed to be TH_p and the letter N is inserted in the appropriate cell in Section 4. Thus (*Fig. 23d*):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P					L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M			Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*d*.

(10) It is about time to try out these assumed values in the message. The proper insertions are made, with the following results:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	HF	CA	PG	OQ	IL	BS	PK	MN	DU	KE	OH	QN	FB	OR	UN
B.	QC	LC	HQ	BQ	BF	HM	AF	XS	IO	KO	QY	FN	SX	MC	GY
C.	XI	FB	EX	AF	DX	LP	MX	HH	RG	KG	QK	QM ST	LF OP	EQ	QI
D.	GO	IH	MU	EO	RD	CL	TU	FE	QQ	CG	QN	HF	XI	FB	EX
E.	FL	BU	QF	CH	QO	QM ST	AF	TX	SY	CB	EP	FN	BS	PK	NU
F.	QI	TX	EU	QM ST	LF OP	EQ	QI	GO	IE	UE	HP	IA	NY	TF	LB
G.	FE	EP	ID	HP	CG	NQ	IH	BF	HM	HF	XC	KU	PD	GQ	PN TH
H.	CB	CQ	LQ	PN TH	FN	PN TH	IT	OR	TE	NC	CB	CN	TF	HH	AY
J.	ZL	QC	IA	AI	QU	CH	TP	CB	IF	GW	KF	CQ	SL	QM ST	CB
K.	OY	CR	QQ	DP	RX	FN	QM ST	LF OP	ID	GC	CG	IO	GO	IH	HF
L.	IR	CG	GG	ND	LN	OZ	TF	GE	ER	RP	IF	HO	TF	HH	AY
M.	ZL	QC	IA	AI	QU	CH	TP								

(11) So far no impossible combinations are in evidence. Beginning with group H4 in the message is seen the following sequence:

P N F N P N
T H . . T H

Assume it to be THAT THE. Then $AT_p = FN_c$, and the letter N is to be inserted in row 4 column 1. But this is inconsistent with previous assumptions, since N in Section 4 has already been tentatively placed in row 2 column 4 of Section 4. Other assumptions for FN_c are made: that it is, IS_p (THIS TH. . .); that it is EN_p (THEN TH. . .); but the same inconsistency is apparent. In fact the student will see that FN_c must represent a digraph ending in F, G, H, I–J, or K, since N_c is tentatively located on the same line as these letters in Section 2. Now FN_c occurs 4 times in the message. The digraph it represents *must* be one of the following:

DF, DG, DH, DI, DJ, DK
IF, IG, IH, II, IJ, IK
JF, JG, JH, JI, JJ, JK
OF, OG, OH, OI, OJ, OK
TK,
YF, YG, YH, YI, YJ, YK

Of these the only one likely to be repeated 4 times is OF, yielding T H O F T H which may be
P N F N P N
a part of

. N O R T H O F T H E . or . S O U T H O F T H E .
C Q L Q P N F N P N I T C Q L Q P N F N P N I T

In either case, the position of the F in Section 3 is excellent: F . . . L in row 3. There are 3 cells intervening between F and L, into which G, H, I–J, and K may be inserted. It is not nearly so likely that G, H, and K are in the keyword as that I should be in it. Let it be assumed that this is the case, and let the letters be placed in the appropriate cells in Section 3. Thus (Fig. 23*e*):

	A	B	C	D	E						
	F	G	H	I–J	K						
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4				F		L	M	N	O	P	2
			M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*e*.

Let the resultant derived values be checked against the frequency distribution. If the position of H in Section 3 is correct, then the digraph ON_p, normally of high frequency should be represented several times by HF_e. Reference to Fig. 22 shows a frequency of 4 times. And HM_e, with 2 occurrences, represents NS_p. There is no need to go through all the possible corroborations.

(12) Going back to the assumption that T H . . T H
P N F N P N
is part of the expression

. N O R T H O F T H E . or . S O U T H O F T H E . .
C Q L Q P N F N P N I T C Q L Q P N F N P N I T

it is seen at once from Fig. 23*e* that the latter is apparently correct and not the former, because LQ_e equals OU_p and not OR_p. If $\Theta S_p = CQ_e$, this means that the letter C of the digraph CQ_e must be placed in row 1 column 3 or row 2 column 3 of Section 3. Now the digraph CB_e occurs 5 times, CG_e, 4 times, CH_e, 3 times, CQ_e, 2 times. Let an attempt be made to deduce the exact position of C in Section 3 and the positions of B, G, and H in Section 4. Since F is already placed in Section

4, assume G and H directly follow it, and that B comes before it. How much before? Suppose a trial be made. Thus (Fig. 23*f*):

	A	B	C	D	E			C			
	F	G	H	I–J	K			C			
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4	B	B	B	F	G	L	M	N	O	P	2
	H		M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*f*.

By referring now to the frequency distribution, Fig. 22, after a very few minutes of experimentation it becomes apparent that the following is correct:

	A	B	C	D	E			C			
	F	G	H	I–J	K						
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N O	P	Q		
	V	W	X	Y	Z						
						A	B	C	D	E	
				N		F	G	H	I–J	K	
4	B			F	G	L	M	N	O	P	2
	H		M	Q		Q	R	S	T	U	
						V	W	X	Y	Z	

FIGURE 23*g*.

(13) The identifications given by these placements are inserted in the text, and solution is very rapidly completed. The final checkerboard and deciphered text are given below.

	A	B	C	D	E	S	O	C	I	E	
	F	G	H	I–J	K	T	Y	A	B	D	
1	L	M	N	O	P	F	G	H	K	L	3
	Q	R	S	T	U	M	N	P	Q	R	
	V	W	X	Y	Z	U	V	W	X	Z	
	E	X	P	U	L	A	B	C	D	E	
	S	I	O	N	A	F	G	H	I–J	K	
	B	C	D	F	G	L	M	N	O	P	
4	H	K	M	Q	R	Q	R	S	T	U	2
	T	V	W	Y	Z	V	W	X	Y	Z	

FIGURE 23A.

A. H F C A P G O Q I L B S P K M N D U K E O H Q N F B O R U N
O N E H U N D R E D F I R S T F I E L D A R T I L L E R Y F

B. Q C L C H Q B Q B F H M A F X S I O K O Q Y F N S X M C G Y
R O M P O S I T I O N S I N V I C I N I T Y O F B A R L O W

C. X I F B E X A F D X L P M X H H R G K G Q K Q M L F E Q Q I
W I L L B E I N G E N E R A L S U P P O R T S T O P D U R I

D. G O I H M U E O R D C L T U F E Q Q C G Q N H F X I F B E X
N G A T T A C K S P E C I A L A T T E N T I O N W I L L B E

E. F L B U Q F C H Q O Q M A F T X S Y C B E P F N B S P K N U
P A I D T O A S S I S T I N G A D V A N C E O F F I R S T B

F. Q I T X E U Q M L F E Q Q I G O I E U E H P I A N Y T F L B
R I G A D E S T O P D U R I N G A D V A N C E I T W I L L P

G. F E E P I D H P C G N Q I H B F H M H F X C K U P D G Q P N
L A C E C O N C E N T R A T I O N S O N W O O D S N O R T H

H. C B C Q L Q P N F N P N I T O R T E N C C B C N T F H H A Y
A N D S O U T H O F T H A Y E R F A R M A N D H I L L S I X

J. Z L Q C I A A I Q U C H T P C B I F G W K F C Q S L Q M C B
Z E R O E I G H T D A S H A A N D O N W O O D S E A S T A N

K. O Y C R Q Q D P R X F N Q M L F I D G C C G I O G O I H H F
D W E S T T H E R E O F S T O P C O M M E N C I N G A T O N

L. I R C G G G N D L N O Z T F G E E R R P I F H O T F H H A Y
E T E N P M S M O K E W I L L B E U S E D O N H I L L S I X

M. Z L Q C I A A I Q U C H T P
Z E R O E I G H T D A S H A

d. (1) It is interesting to note how much simpler the matter becomes when the positions of the plain-text and cipher-text sections are reversed, or, what amounts to the same thing, when in encipherment the plain-text pairs are sought in the sections containing the mixed alphabets, and their cipher equivalents are taken from the sections containing the normal alphabets. For example, referring to Fig. 23*h*, suppose that sections 3–4 be used as the source of the plain-text pairs, and sections 1–2 as the source of the cipher-text pairs. Then $ON_b = DG_a$, $EH_b = AU_a$, etc.

(2) To solve a message enciphered in that manner, it is necessary merely to make a square in which all four sections are normal alphabets, and then perform two steps. First, the cipher text pairs are converted into their normal alphabet equivalents merely by "deciphering" the message with that square; the result of this operation yields two monoalphabets, one composed of the odd letters, the other of the even letters. The second step is to solve these two mono-alphabets.

(3) Where the same mixed alphabet is inserted in sections 3 and 4, the problem is still easier, since the letters resulting from the conversion into normal-alphabet equivalents all belong to the same, single-mixed alphabet.

45. Analysis of ciphers based upon other types of checkerboard designs.—The solution of cryptograms enciphered by other types of checkerboard designs is accomplished along lines very similar to those set forth in the foregoing example of the solution of a message prepared by means of a 4-square checkerboard design. There are, unfortunately, no means or tests which can be applied to determine in the early stages of the analysis exactly what type of design is involved in the *first* case under study. The author freely admits that the solution outlined in subparagraph *c* is quite artificial in that nothing is demonstrated in step (7) that obviously leads to or warrants the assumption that a 4-square checkerboard is involved. This point was passed over with the quite bald statement that this was "the simplest thing to assume"—and then the solution proceeds exactly as though this mere *hypothesis* has been definitely established. For example, the very first results obtained were based upon assuming that a certain 4-letter repetition represented the word STOP and *immediately inserting certain letters in appropriate cells in a 4-square checkerboard.* Several more assumptions were built on top of that and very rapid strides were made. What if it had not been a 4-square checkerboard at all? What if it had been a 2-square checkerboard of the type shown in Fig. 24?

M	A	N	U	F	O	S	Q	L	P
C	T	R	I	G	W	Z	Y	V	X
B	D	E	H	K	D	K	H	B	E
L	O	P	Q	S	A	F	U	M	N
V	W	X	Y	Z	T	G	I	C	R

FIGURE 24.

The only defense that can be made of what may seem to the student to be purely arbitrary procedure based upon the author's advance information or knowledge is the following: In the first place, in order to avoid making the explanation a too-long-drawn-out affair, it is necessary (and pedagogical experience warrants) that certain alternative hypotheses be passed over in silence. In the second place, it may now be added, *after* the principles and procedure have been elucidated (which at this stage is the primary object of this text) that if good results do not follow from a first hypothesis, the only thing the cryptanalyst can do is to reject that hypothesis, and formulate a second hypothesis. In actual practice he may have to reject a second, third, fourth, . . . *nth* hypothesis. In the end he may strike the right one—or he may not. There is no guaranty of success in the matter. In the third place, one of the objects of this text is to show how certain systems, if employed for military purposes, can readily be broken down. Assuming

that a checkerboard system is in use, and that daily changes in keywords are made, it is possible that the traffic of the first day might give considerable difficulty in solution, if the type of checkerboard were not known to the cryptanalyst. But the second or third day's traffic would be easy to solve, because by that time the cryptanalytic personnel would have analyzed the system and thus learned what type of checkerboard the enemy is using.

46. Analysis of the Playfair cipher system.—*a.* An excellent example of a practical, partially digraphic system is the Playfair cipher.[6] It was used for a number of years as a field cipher by the British Army, before and during the World War, and for a short time, also during that war, by certain units of the American Expeditionary Forces.

b. Published solutions[7] for this cipher are quite similar basically and vary only in minor details. The earliest, that by Lieut. Mauborgne, used straightforward principles of frequency to establish the values of three or four of the most frequent digraphs. Then, on the assumption that in most cases in which a keyword appears on the first and second rows the last five letters of the normal alphabet, VWXYZ, will rarely be disturbed in sequence and will occupy the last row of the square, he "juggles" the letters given by the values tentatively established from frequency considerations, placing them in various positions in the square, together with VWXYZ, to correspond to the plain-text cipher relationships tentatively established. A later solution by Lieut. Frank Moorman, as described in Hitt's Manual, assumes that in a Playfair cipher prepared by means of a square in which the keyword occupies the first and second rows, if a digraphic frequency distribution is made, it will be found that the letters having the greatest combining power are very probably letters of the key. A still later solution, by Lieut. Commander Smith, is perhaps the most lucid and systematized of the three. He sets forth in definite language certain considerations which the other two writers certainly entertained but failed to indicate.

c. The following details have been summarized from Commander Smith's solution:

(1) The Playfair cipher may be recognized by virtue of the fact that it always contains an even number of letters, and that when divided into groups of two letters each, no group contains a repetition of the same letter, as NN or EE. Repetitions of digraphs, trigraphs, and polygraphs will be evident in fairly long messages.

(2) Using the square[8] shown in Fig. 25*a*, there are two general cases to be considered, as regards the results of encipherment:

B	A	N	K	R
D	E	F	G	H
I–J	L	M	O	Q
U	P	T	C	Y
S	V	W	X	Z

FIGURE 25*a*.

[6] This cipher was really invented by Sir Charles Wheatstone but receives its name from Lord Playfair, who apparently was its sponsor before the British Foreign Office. See Wemyss Reid, *Memoirs of Lyon Playfair*, London, 1899. A detailed description of this cipher will be found in Sec. VIII, *Advanced Military Cryptography*.

[7] Mauborgne, Lieut. J. O., U. S. A. *An advanced problem in cryptography and its solution*, Leavenworth, 1914.

Hitt, Captain Parker, U. S. A. *Manual for the solution of military ciphers*, Leavenworth, 1918.

Smith, Lieut. Commander W. W., U. S. N. In *Cryptography* by André Langie, translated by J. C. H. Macbeth, New York, 1922.

[8] The Playfair square accompanying Commander Smith's solution is based upon the keyword BANKRUPTCY, "to be distributed between the first and fourth lines of the square." This is a simple departure from the original Playfair scheme in which the letters of the keyword are written from left to right and in consecutive lines from the top downward.

CASE 1. Letters at opposite corners of a rectangle. The following illustrative relationships are found:

$$TH_p=YF_c$$
$$HT_p=FY_c$$
$$YF_p=TH_c$$
$$FY_p=HT_c$$

Reciprocity is complete.

CASE 2. Two letters in the same line or column. The following illustrative relationships are found:

$$AN_p=NK_c$$
$$NA_p=KN_c$$

But NK_p does not $=AN_c$, nor does $KN_p=NA_c$.

Reciprocity is only partial.

(3) The foregoing gives rise to the following:

RULE I. (*a*) Regardless of the position of the letters in the square, if

$$1.2=3.4, \text{ then}$$
$$2.1=4.3$$

(*b*) If 1 and 2 form opposite corners of a rectangle, the following equations obtain:

$$1.2=3.4$$
$$2.1=4.3$$
$$3.4=1.2$$
$$4.3=2.1$$

(4) A letter considered as occupying a position in a row can be combined with but four other letters in the same row; the same letter considered as occupying a position in a column can be combined with but four other letters in the same column. Thus, this letter can be combined with only 8 other letters all told, under Case 2, above. But the same letter considered as occupying a corner of a rectangle can be combined with 16 other letters, under Case 1, above. Commander Smith derives from these facts the conclusion that "it would appear that Case 1 is twice as probable as Case 2." He continues thus (notation my own):

"Now in the square, note that:

$AN_p=NK_c$		$EN_p=FA_c$
$GN_p=FK_c$		$EM_p=FL_c$
$ON_p=MK_c$	also	$ET_p=FP_c$
$CN_p=TK_c$		$EW_p=FV_c$
$XN_p=WK_c$		$EF_p=FG_c$

"From this it is seen that of the 24 equations that can be formed when each letter of the square is employed either as the initial or final letter of the group, five will indicate a repetition of a corresponding letter of plain text.

"Hence, RULE II. After it has been determined, in the equation $1.2=3.4$, that, say, $EN_p=FA_c$, there is a probability of one in five that any other group beginning with F_c indicates $E\Theta_p$, and that any group ending in A_c indicates ΘN_p.

"After such combinations as ER_p, OR_p, and EN_p have been assumed or determined, the above rule may be of use in discovering additional digraphs and partial words." [9]

RULE III. In the equation $1.2=3.4$, 1 and 3 can never be identical, nor can 2 and 4 ever be identical. Thus, AN_p could not possibly be represented by AY_c, nor could ER_p be represented by KR_c. This rule is useful in elimination of certain possibilities when a specific message is being studied.

RULE IV. In the equation $1.2_p=3.4_c$, if 2 and 3 are identical, the letters are all in the same row or column, and in the relative order 124. In the square shown, $AN_p=NK_c$ and the absolute order is ANK. The relative order 124 includes five absolute orders which are cyclic permutations of one another. Thus: ANK. . , NK. . A, K. . AN, . . ANK, and . ANK. .

RULE V. In the equation $1.2_p=3.4_c$, if 1 and 4 are identical, the letters are all in the same row or column, and in the relative order 243. In the square shown, $KN_p=RK_c$ and the absolute order is NKR. The relative order 243 includes five absolute orders which are cyclic permutations of one another. Thus NKR. . , KR. . N, R. . NK, . . NKR, and . NKR. .

RULE VI. "Analyze the message for group recurrences. Select the groups of greatest recurrence and assume them to be high-frequency digraphs. Substitute the assumed digraphs throughout the message, testing the assumptions in their relation to other groups of the cipher. The reconstruction of the square proceeds simultaneously with the solution of the message and aids in hastening the translation of the cipher."

d. (1) When solutions for the Playfair cipher system were first developed, based upon the fact that the letters were inserted in the cells in keyword-mixed order, cryptographers thought it desirable to place stumbling blocks in the path of such solution by departing from strict, keyword-mixed order. Playfair squares of the latter type are designed as "modified Playfair squares." One of the simplest methods is illustrated in Fig. 25*a*, wherein it will be noted that the last five letters of the keyword proper are inserted in the fourth row of the square instead of the second, where they would naturally fall. Another method is to insert the letters within the cells from left to right and top downward but use a sequence that is a keyword-mixed sequence developed by a columnar transposition based upon the keyword proper. Thus, using the keyword BANKRUPTCY:

2 1 5 4 7 9 6 8 3 10
B A N K R U P T C Y
D E F G H I L M O Q
S V W X Z

Sequence: A E V B D S C O K G X N F W P L R H Z T M U I Y Q

[9] There is an error in this reasoning. Take, for example, the 24 equations having F as an initial letter:

Case		Case		Case		Case	
1.	$FB_c=DN_p$	2.	FE=ED	2.	FT=NM	1.	FX=GW
2.	FD =EH	1.	FL=EM	2.	FW=NT	1.	FR=HN
1.	FI =DM	1.	FP=ET	1.	FK=GN	2.	FH=EG
1.	FU =DT	1.	FV=EW	2.	FG=EF	1.	FQ=HM
1.	FS =DW	2.	FN=NW	1.	FO=GM	1.	FY=HT
1.	FA =EN	2.	FM=NF	1.	FC=GT	1.	FZ=HW

Here, the initial letter F_c represents the following initial letters of plain-text digraphs:

D E N G H

It is seen that F_c represents D_p, N_p, G_p, H_p 4 times each, and E_p, 8 times. Consequently, supposing that it has been determined that $FA_c=EN_p$, the probability that F_c will represent E_p is not 1 in 5 but 8 in 24, or 1 in 3; but supposing that it has been determined that $FW_c=NT_p$, the probability that F_c will represent N_p is 4 in 24 or 1 in 6. The difference in these probabilities is occasioned by the fact that the first instance, $FA_c=EN_p$ corresponds to a Case 1 encipherment, the second instance, $FW_c=NT_p$, to a Case 2 encipherment. But there is no way of knowing initially, and without other data, whether one is dealing with a Case 1 or Case 2 encipherment. Only as an approximation, therefore, may one say that the probability of F_c representing a given Θ_p is 1 in 5.

The Playfair Square is as follows:

A	E	V	B	D
S	C	O	K	G
X	N	F	W	P
L	R	H	Z	T
M	U	I	Y	Q

FIGURE 25b.

(2) In the foregoing square practically all indications that the square has been developed from a keyword have disappeared. The principal disadvantage of such an arrangement is that it requires more time to locate the letters desired, both in cryptographing and decryptographing, than it usually does when a semblance of normal alphabetic order is preserved in the square.

(3) Note the following three squares:

Z	T	L	R	H
Y	Q	M	U	I
B	D	A	E	V
K	G	S	C	O
W	P	X	N	F

FIGURE 25c.

O	K	G	S	C
H	Z	T	L	R
V	B	D	A	E
F	W	P	X	N
I	Y	Q	M	U

FIGURE 25d.

N	F	W	P	X
R	H	Z	T	L
U	I	Y	Q	M
E	V	B	D	A
C	O	K	G	S

FIGURE 25e.

At first glance they all appear to be different, but closer examination shows them to be *cyclic permutations* of one another and of the square in Fig. 25*b*. They yield identical equivalents in all cases. However, if an attempt be made to reconstruct the original keyword, it would be much easier to do so from Fig. 25*b* than from any of the others, because in Fig. 25*b* the keyword-mixed sequence has not been disturbed as much as in Figs. 25*c*, *d*, *e*. In working with Playfair ciphers, the student should be on the lookout for such instances of cyclic permutation of the original Playfair square, for during the course of solution he will not know whether he is building up the original or an equivalent cyclic permutation of the original square; only after he has completely reconstructed the square will he be able to determine this point.

(4) It can readily be shown that the columns of a Playfair square may be cyclically permuted (see subpar. *d*) to produce a first set of 25 squares all of which, though at first glance apparently different, will yield identical equivalents; likewise, the rows of such a square may be cyclically permuted to produce a second set of 25 squares all of which will also yield identical equivalents. Thus there may be a total of 50 cyclic permutations composed of two sets of 25 each. The cipher equivalents yielded by Case 2 encipherments (letters in the same row or in the same column) will be identical for any two of these 50 different Playfair squares; but the cipher equivalents yielded by Case 1 encipherments (letters at diagonally opposite corners of a rectangle) will be identical only for two squares belonging to the same set of 25 cyclic permutations.

e. (1) The steps in the solution of a typical example of this cipher may be useful. Let the message be as follows:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
A.	V	T	Q	E	U	H	I	O	F	T	C	H	X	S	C	A	K	T	V	T	R	A	Z	E	V	T	A	G	A	E
B.	O	X	T	Y	M	H	C	R	L	Z	Z	T	Q	T	D	U	M	C	Y	C	X	C	T	G	M	T	Y	C	Z	U
C.	S	N	O	P	D	G	X	V	X	S	C	A	K	T	V	T	P	K	P	U	T	Z	P	T	W	Z	F	N	B	G
D.	P	T	R	K	X	I	X	B	P	R	Z	O	E	P	U	T	O	L	Z	E	K	T	T	C	S	N	H	C	Q	M
E.	V	T	R	K	M	W	C	F	Z	U	B	H	T	V	Y	A	B	G	I	P	R	Z	K	P	C	Q	F	N	L	V
F.	O	X	O	T	U	Z	F	A	C	X	X	C	P	Z	X	H	C	Y	N	O	T	Y	O	L	G	X	X	I	I	H
G.	T	M	S	M	X	C	P	T	O	T	C	X	O	T	T	C	Y	A	T	E	X	H	F	A	C	X	X	C	P	Z
H.	X	H	Y	C	T	X	W	L	Z	T	S	G	P	Z	T	V	Y	W	C	E	T	W	G	C	C	M	B	H	M	Q
J.	Y	X	Z	P	W	G	R	T	I	V	U	X	P	U	M	Q	R	K	M	W	C	X	T	M	R	S	W	G	H	B
K.	X	C	P	T	O	T	C	X	O	T	M	I	P	Y	D	N	F	G	K	I	T	C	O	L	X	U	E	T	P	X
L.	X	F	S	R	S	U	Z	T	D	B	H	O	Z	I	G	X	R	K	I	X	Z	P	P	V	Z	I	D	U	H	Q
M.	O	T	K	T	K	C	C	H	X	X																				

(2) Without going through the preliminary tests in detail, with which it will be assumed that the student is now familiar,[10] the conclusion is reached that the cryptogram is digraphic in nature, and a digraphic frequency distribution is made (Fig. 26).

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A					1		1																		
B							2	2																	
C	2				1	1		2				1				1	1						5	1	
D		1					1						1							2					
E															1				1						
F	2						1						2						1						
G			1																				2		
H		1	1											1		1									
I								1						1	1						1		1		
K			1						1						1				4						
L																					1				1
M			1					1	1							2			1			2			
N														1											
O											3				1				6				2		
P										1							1		4	2	1		1	1	3
Q					1							1							1						
R	1									4								1	1						1
S							1					1	2				1			1					
T			3		1		1					2									2	1	1	2	1
U								1											1				1		1
V																			5						
W							2				1														1
X		1	5			1		3	2									2		1	1				
Y	2		3																			1	1		
Z					2				2					1	2				3	2					

FIGURE 26.

[10] See Par. 44c.

Since there are no double-letter groups, the conclusion is reached that a Playfair cipher is involved and the message is rewritten in digraphs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A.	VT	QE	UH	IO	FT	CH	XS	CA	KT	VT	RA	ZE	VT	AG	AE
B.	OX	TY	MH	CR	LZ	ZT	QT	DU	MC	YC	XC	TG	MT	YC	ZU
C.	SN	OP	DG	XV	XS	CA	KT	VT	PK	PU	TZ	PT	WZ	FN	BG
D.	PT	RK	XI	XB	PR	ZO	EP	UT	OL	ZE	KT	TC	SN	HC	QM
E.	VT	RK	MW	CF	ZU	BH	TV	YA	BG	IP	RZ	KP	CQ	FN	LV
F.	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY	OL	GX	XI	IH
G.	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ
H.	XH	YC	TX	WL	ZT	SG	PZ	TV	YW	CE	TW	GC	CM	BH	MQ
J.	YX	ZP	WG	RT	IV	UX	PU	MQ	RK	MW	CX	TM	RS	WG	HB
K.	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET	PX
L.	XF	SR	SU	ZT	DB	HO	ZI	GX	RK	IX	ZP	PV	ZI	DU	HQ
M.	OT	KT	KC	CH	XX										

(3) The following three fairly lengthy repetitions are noted:

Lines									
F.	OT	UZ	FA	CX	XC	PZ	XH	CY	NO
G.	TE	XH	FA	CX	XC	PZ	XH	YC	TX
A.	FT	CH	XS	CA	KT	VT	RA	ZE	
C.	DG	XV	XS	CA	KT	VT	PK	PU	
G.	TM	SM	XC	PT	OT	CX	OT	TC	
K.	WG	HB	XC	PT	OT	CX	OT	MI	

The first long repetition, with the sequent reversed digraphs CX and XC immediately suggests the word BATTALION, split up into –B AT TA LI ON and the sequence containing this repetition in lines F and G becomes as follows:

Line F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
				B	AT	TA	LI	ON			
Line G	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
			ON	B	AT	TA	LI	ON			

(4) Because of the frequent use of numerals before the word BATTALION and because of the appearance of ON before this word in line G, the possibility suggests itself that the word before BATTALION in line G is either ONE or SECOND. The identical digraph FA in both cases gives a hint that the word BATTALION in line F may also be preceded by a numeral; if ONE is

correct in line G, then THREE is possible in line F. On the other hand, if SECOND is correct in line G, then THIRD is possible in line F. Thus:

Line F	OX	OT	UZ	FA	CX	XC	PZ	XH	CY	NO	TY
1st hypothesis	—	TH	RE	EB	AT	TA	LI	ON			
2nd hypothesis	—	TH	IR	DB	AT	TA	LI	ON			
Line G	YA	TE	XH	FA	CX	XC	PZ	XH	YC	TX	WL
1st hypothesis	—	—	ON	EB	AT	TA	LI	ON			
2nd hypothesis	–S	EC	ON	DB	AT	TA	LI	ON			

First, note that if either hypothesis is true, then $OT_c=TH_p$. The frequency distribution shows that OT occurs 6 times and is in fact the most frequent digraph in the message. Moreover, by Rule I of subparagraph *b*, if $OT_c=TH_p$ then $TO_c=HT_p$. Since HT_p is a very rare digraph in normal plain text, TO_c should either not occur at all in so short a message or else it should be very infrequent. The frequency distribution shows its entire absence. Hence, there is nothing inconsistent with the possibility that the word in front of BATTALION in line F is THREE or THIRD, and some evidence that it is actually one or the other.

(5) But can evidence be found for the support of one hypothesis against the other? Let the frequency distribution be examined with a view to throwing light upon this point. If the first hypothesis is true, then $UZ_c=RE_p$, and, by Rule I, $ZU_c=ER_p$. The frequency distribution shows but one occurrence of UZ_c and but two occurrences of ZU_c. These do not look very good for RE and ER. On the other hand, if the second hypothesis is true, then $UZ_c=IR_p$ and, by Rule I, $ZU_c=RI_p$. The frequencies are much more favorable in this case. Is there anything inconsistent with the assumption, on the basis of the second hypothesis, that $TE_c=EC_p$? The frequency distribution shows no inconsistency, for TE_c occurs once and ET_c ($=CE_p$, by Rule I) occurs once. As regards whether $FA_c=EB_p$ or DB_p, both hypotheses are tenable; possibly the second hypothesis is a shade better than the first, on the following reasoning: By Rule I, if $FA_c=EB_p$ then $AF_c=BE_p$, or if $FA_c=DB_p$ then $AF_c=BD_p$. The fact that no AF_c occurs, whereas at least one BE_p may be expected in this message, inclines one to the second hypothesis, since BD_p is very rare.

(6) Let the 2nd hypothesis be assumed to be correct. The additional values are tentatively inserted in the text, and in lines G and K two interesting repetitions are noted:

Line G	TM	SM	XC	PT	OT	CX	OT	TC	YA	TE	XH	FA	CX	XC	PZ	XH
			TA		TH	AT	TH		–S	EC	ON	DB	AT	TA	LI	ON
Line K	WG	HB	XC	PT	OT	CX	OT	MI	PY	DN	FG	KI	TC	OL	XU	ET
			TA		TH	AT	TH									

This certainly looks like STATE THAT THE. . ., which would make $TE_p=PT_c$. Furthermore, in line G the sequence STATETHATTHE..SECONDBATTALION can hardly be anything else than STATE THAT THEIR SECOND BATTALION, which would make $TC_c=EI_p$ and $YA_c=RS_p$. Also $SM_c=-S_p$.

(7) It is perhaps high time that the whole list of tentative equivalent values be studied in relation to their consistency with the positions of letters in the Playfair square; moreover, by so doing, additional values may be obtained in the process. The complete list of values is as follows:

Assumed values	*Derived by Rule I*
$AT_p=CX_e$	$TA_p=XC_e$
$LI_p=PZ_e$	$IL_p=ZP_e$
$ON_p=XH_e$	$NO_p=HX_e$
$TH_p=OT_e$	$HT_p=TO_e$
$IR_p=UZ_e$	$RI_p=ZU_e$
$DB_p=FA_e$	$BD_p=AF_e$
$EC_p=TE_e$	$CE_p=ET_e$
$TE_p=PT_e$	$ET_p=TP_e$
$EI_p=TC_e$	$IE_p=CT_e$
$RS_p=YA_e$	$SR_p=AY_e$
$-S_p=SM_e$	$S-_p=MS_e$

(8) By Rule V, the equation $TH_p=OT_e$ means that H, T, and O are all in the same row or column and in the relative order 2–4–3; similarly, C, E, and T are in the same row or column and in the relative order 243. Further E, P, and T are in the same row and column, and their relative order is also 243. That is, these sequences must occur in the square:

(1)	(2)	(3)
H T O . . , or	C E T . . , or	E T P . . , or
T O . . H , or	E T . . C , or	T P . . E , or
O . . H T , or	T . . C E , or	P . . E T , or
. . H T O , or	. . C E T , or	. . E T P , or
. H T O .	. C E T .	. E T P .

(9) Noting the common letters E and T in the second and third sets of relative orders, these may be combined into one sequence of four letters. Only one position remains to be filled and noting, in the list of equivalents that $EI_p=TC_e$, it is obvious that the letter I belongs to the CET sequence. The complete sequence is therefore as follows:

C E T P I , or
E T P I C , or
T P I C E , or
P I C E T , or
I C E T P

(10) Taking up the HTO sequence, it is noted, in the list of equivalents that $ON_p=XH_e$, an equation containing two of the three letters of the HTO sequence. From this it follows that N and X must belong to the same row or column as HTO. The arrangement must be one of the following:

H T O X N
T O X N H
O X N H T
X N H T O
N H T O X

(11) Since the sequence containing HTOXN has a common letter (T) with the sequence CETPI, it follows that if the HTOXN sequence occupies a row, then the CETPI sequence must occupy a column; or, if the HTO sequence occupies a column, then the CETPI sequence must

occupy a row; and they may be combined by means of their common letter, T. According to subpar. *d* (4), the two sequences may be inserted within a Playfair square in 25 different ways by cyclically permuting and shifting the letters of one of these two sequences; and the same two sequences may be again inserted in another set of 25 ways by cyclically permuting and shifting the letters of the other of these two sequences. In Fig. 27 the diagrams labeled (1) to (10), inclusive, show 10 of the possible 25 obtainable by making the HTOXN sequence one of the rows of the square; diagrams (11) and (12) show 2 of the possible 25 obtainable by making the HTOXN sequence one of the columns of the square. The entire complement of 25 arrangements for each set may easily be drawn up by the student; space forbids their being completely set forth and it is really unnecessary to do so.

(1)

C				
E				
T	O	X	N	H
P				
I				

(2)

	C			
	E			
H	T	O	X	N
	P			
	I			

(3)

		C		
		E		
N	H	T	O	X
		P		
		I		

(4)

			C	
			E	
X	N	H	T	O
			P	
			I	

(5)

				C
				E
O	X	N	H	T
				P
				I

(6)

P				
I				
C				
E				
T	O	X	N	H

(7)

	P			
	I			
	C			
	E			
H	T	O	X	N

(8)

		P		
		I		
		C		
		E		
N	H	T	O	X

(9)

			P	
			I	
			C	
			E	
X	N	H	T	O

(10)

				P
				I
				C
				E
O	X	N	H	T

(11)

				N
				H
P	I	C	E	T
				O
				X

(12)

N				
H				
T	P	I	C	E
O				
X				

FIGURE 27.

(12) Before trying to discover means whereby the actual or absolute arrangement may be detected from among the full set of 50 possible arrangements, the question may be raised: is it necessary? So far as concerns Case 2 encipherments, since any one of the 50 arrangements will yield the same equivalents as any of the remaining 49, perhaps a relative arrangement will do.

(13) Let arrangement 8 be arbitrarily selected for trial.

		P		
		I		
		C		
		E		
N	H	T	O	X

FIGURE 28a.

(14) What additional letters can be inserted, using as a guide the list of equivalents in subparagraph (7)? There is $AT_p=CX_c$, for example. It contains only one letter, A, not in the arrangement selected for trial, and this letter may immediately be placed, as shown:[10]

		P		
		I		
		C		A
		E		
N	H	T	O	X

FIGURE 28b.

Scanning the list for additional cases of this type, none are found. But seeing that several high-frequency letters have already been inserted in the square, perhaps reference to the cryptogram itself in connection with values derived from these inserted letters may yield further clues. For example, the vowels A, E, I, and O are all in position, as are the very frequent consonants N and T. The following combinations may be studied:

$AN_p=OX_c$	$AT_p=CX_c$	$NA_p=XO_c$	$TA_p=XC_c$
$EN_p=OT_c$	$ET_p=TP_c$	$NE_p=TO_c$	$TE_p=PT_c$
$IN_p=OT_c$	$IT_p=CP_c$	$NI_p=TO_c$	$TI_p=PC_c$
$ON_p=XH_c$	$OT_p=XO_c$	$NO_p=HX_c$	$TO_p=OX_c$

$AT_p(=CX_c)$, $TA_p(=XC_c)$, $ON_p(=XH_c)$, $TE_p(=PT_c)$ and $ET_p(=TP_c)$ have already been inserted in the text. Of the others, only $OX_c(=TO_p)$ occurs two times, and this value can be at once inserted in the text. But can the equivalents of AN, EN, or IN be found from frequency considerations?

[10] The fact that the placement of A yields $AT_p=CX_c$ means that the outline selected for experiment really belongs to the correct set of 25 possible cyclic permutations, and that the letters of the NHTOX sequence belong in a row, the letters of the PICET sequence belong in a column of the original Playfair square. If the reverse were the case, one could not obtain $AT_p=CX_c$ but would obtain $AT_p=XC_c$.

Take EN_p, for example; it is represented by ΘT_c. What combination of ΘT is most likely to represent EN_p among the following candidates:

KT_c (4 times); by Rule I, NE_p would $= TK_c$ (no occurrences)

VT_c (5 times); by Rule I, NE_p would $= TV_c$ (2 times)

ZT_c (3 times); by Rule I, NE_p would $= TZ_c$ (1 time)

VT_c certainly looks good: it begins the message, suggesting the word ENEMY; in line H, in the sequence PZTV would become LINE. Let this be assumed to be correct, and let the word ENEMY also be assumed to be correct. Then $EM_p = QE_c$ and the square then becomes as shown herewith:

		P		
		I		
		C		A
V	M	E	Q	
N	H	T	O	X

FIGURE 28c.

(15) In line E is seen the following sequence:

Line E: . . . . . VT RK MW CF ZU BH TV YA BG IP RZ KP CQ FN LV
EN RI NE RS PT E

The sequence . . .RI..NERS..PT. . . suggests PRISONERS CAPTURED, as follows:

MW CF ZU BH TV YA BG IP RZ KP
P RI SO NE RS CA PT UR ED

This gives the following new values: $\Theta P_p = CF_c$; $SO_p = BH_c$; $CA_p = BG_c$; $UR_p = RZ_c$; $ED_p = KP_c$.

The letters B and G can be placed in position at once, since the positions of C and A are already known. The insertion of the letter B immediately permits the placement of the letter S, from the equation $SO_p = BH_c$. Of the remaining equations only $ED_p = KP_c$ can be used. Since E and P are fixed and are in the same column, D and K must be in the same column, and moreover the K must be in the same row as E. There is only one possible position for K, viz, immediately after Q. This automatically fixes the position of D. The square is now as shown herewith:

		P		D
		I		
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

FIGURE 28d.

(16) A review of all equations, including the very first ones established, gives the following which may now be used: $DB_p=FA_e$; $RS_p=YA_e$. The first permits the immediate placement of F; the second, by elimination of possible positions, permits the placement of both R and Y. The square is now as shown herewith:

		P	F	D
	Y	I		R
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

FIGURE 28e.

Once more a review is made of all remaining thus far unused equations. $LI_p=PZ_e$ now permits the placement of L and Z. $IR_p=UZ_e$ now permits the placement of U, which is confirmed by the equation $UR_p=RZ_e$ from the word CAPTURED.

L		P	F	D
Z	Y	I	U	R
G	S	C	B	A
V	M	E	Q	K
N	H	T	O	X

FIGURE 28f.

There is then only one cell vacant, and it must be occupied by the only letter left unplaced, viz, W. Thus the whole square has been reconstructed, and the message can now be decryptographed.

(17) Is the square just reconstructed identical with the original, or is it a cyclic permutation of a keyword-mixed Playfair square of the type illustrated in Fig. 25*b*? Even though the message can be read with ease, this point is still of interest. Let the sequence be written in five ways, each composed of five partial sequences made by cyclicly permuting each of the horizontal rows of the reconstructed square. Thus:

	Row 1	Row 2	Row 3	Row 4	Row 5
(a)	L W P F D	Z Y I U R	G S C B A	V M E Q K	N H T O X
(b)	W P F D L	Y I U R Z	S C B A G	M E Q K V	H T O X N
(c)	P F D L W	I U R Z Y	C B A G S	E Q K V M	T O X N H
(d)	F D L W P	U R Z Y I	B A G S C	Q K V M E	O X N H T
(e)	D L W P F	R Z Y I U	A G S C B	K V M E Q	X N H T O

By experimenting with these five sequences, in an endeavor to reconstruct a transposition rectangle conformable to a keyword sequence, the last sequence yields the following:

P Y A C M N
D F I G B E H
L R U S K Q T
W Z V X O

By shifting the O from the last position to the first, and rearranging the columns, the following is obtained:

2 5 3 6 1 4 7
C O M P A N Y
B D E F G H I
K L Q R S T U
V W X Z

The original square must have been this:

A	G	S	C	B
K	V	M	E	Q
X	N	H	T	O
D	L	W	P	F
R	Z	Y	I	U

FIGURE 28*g*

f. Continued practice in the solution of Playfair ciphers will make the student quite expert in the matter and will enable him to solve shorter and shorter messages.[11] Also, with practice it will become a matter of indifference to him as to whether the letters are inserted in the square with any sort of regularity, such as simple keyword-mixed order, columnar transposed keyword-mixed order, or in a purely random order.

g. It may perhaps seem to the student that the foregoing steps are somewhat too artificial, a bit too "cut and dried" in their accuracy to portray the process of analysis, as it is applied in practice. For example, the critical student may well object to some of the assumptions and the reasoning in step (5) above, in which the words THREE and ONE (1st hypothesis) were rejected in favor of the words THIRD and SECOND (2nd hypothesis). This rested largely upon the rejection of RE_p and ER_p as the equivalents of UZ_c and ZU_c, and the adoption of IR_p and RI_p as their equivalents. Indeed, if the student will examine the final message with a critical eye he will find that while the bit of reasoning in step (5) is perfectly logical, the assumption upon which it is based is in fact wrong, for it happens that in this case ER_p occurs only once and RE_p does not occur at all. Consequently, although most of the reasoning which led to the rejection of the 1st hypothesis and the adoption of the 2nd was logical, it was in fact based upon erroneous assump-

[11] The author once had a student who "specialized" in Playfair ciphers and became so adept that he could solve messages containing as few as 50–60 letters within 30 minutes.

tion. In other words, despite the fact that the assumption was incorrect, a correct deduction was made. *The student should take note that in cryptanalysis situations of this sort are not at all unusual.* Indeed they are to be expected and a few words of explanation at this point may be useful.

h. Cryptanalysis is a science in which deduction, based upon observational data, plays a very large role. But it is also true that in this science most of the deductions usually rest upon assumptions. It is most often the case that the cryptanalyst is forced to make his assumptions upon a quite limited amount of text. It cannot be expected that assumptions based upon statistical generalizations will always hold true when applied to data comparatively very much smaller in quantity than the total data used to derive the generalized rules. Consequently, as regards assumptions made in specific messages, *most of the time* they will be correct, but *occasionally* they will be incorrect. In cryptanalysis it is often found that among the correct deductions there will be cases in which subsequently discovered facts do not bear out the assumptions on which the deduction was based. Indeed, it is sometimes true that if the *facts* had been known *before* the deduction was made, this knowledge would have prevented making the correct deduction. For example, suppose the cryptanalyst had somehow or other divined that the message under consideration contained no RE, only one ER, one IR, and two RI's (as is actually the case). He would certainly not have been able to choose between the words THREE and ONE (1st hypothesis) as against THIRD and SECOND (2d hypothesis). But because he assumes that there should be more ER_p's and RE_p's than IR's and RI's in the message, he deduces that UZ_e cannot be RE_p, rejects the 1st hypothesis and takes the 2d. It later turns out, after the problem has been solved, that the deduction was correct, although the assumption on which it was based (expectation of more frequent appearance of RE_p and ER_p) was, in fact, *not* true in this particular case. The cryptanalyst can only hope that the number of times when his deductions are correct, even though based upon assumptions which later turn out to be erroneous, will abundantly exceed the number of times when his deductions are wrong, even though based upon assumptions which later prove to be correct. If he is lucky, the making of an assumption which is really not true will make no difference in the end and will not delay solution; but if he is specially favored with luck, it may actually help him solve the message—as was the case in this particular example.

i. Another comment of a general nature may be made in connection with this specific example. The student may ask what would have been the procedure in this case if the message had not contained such a tell-tale repetition as the word BATTALION, which formed the point of departure for the solution, or, as it is often said, permitted an "entering wedge" to be driven into the message. The answer to his query is that if the word BATTALION had not been repeated, there would probably have been some other repetition which would have permitted the same sort of attack. If the student is looking for cut and dried, straight-forward, unvarying methods of attack, he should remember that cryptanalysis, while it may be considered a branch of mathematics, is not a science which has many "general solutions" such as are found and expected in mathematics proper. It is inherent in the very nature of cryptanalytics that, *as a rule*, only general principles can be established; their practical application must take advantage of peculiarities and particular situations which are noted in specific messages. This is especially true in a text on the subject. The illustration of a general principle requires a specific example, and the latter must of necessity manifest characteristics which make it different from any other example. The word BATTALION was not purposely repeated in this example in order to make the demonstration of solution easy; "it just happened that way." In another example, some other entering wedge would have been found. The student can be expected to learn only the *general principles* which will enable him to take advantage of the *specific characteristics* manifested in *specific cases*. Here it is desired to illustrate the general principles of solving Playfair ciphers and to point out the fact that entering wedges must and can be found. The specific nature of the entering wedge varies with specific examples.

SECTION X

CONCLUDING REMARKS

	Paragraph
Special remarks concerning the initial classification of cryptograms	47
Ciphers employing characters other than letters or figures	48
Concluding remarks concerning monoalphabetic substitution	49
Analytical key for cryptanalysis	50

47. Special remarks concerning the initial classification of cryptograms.—*a.* The student should by this time have a good conception of the basic nature of monoalphabetic substitution and of the many "changes" which may be rung upon this simple tune. The first step of all, naturally, is to be able to classify a cryptogram properly and place it in either the transposition or the substitution class. The tests for this classification have been given and as a rule the student will encounter no difficulty in this respect.

b. There are, however, certain kinds of cryptograms whose class cannot be determined in the usual manner, as outlined in Par. 13 of this text. First of all there is the type of code message which employs bona-fide dictionary words as code groups.[1] Naturally, a frequency distribution of such a message will approximate that for normal plain text. The appearance of the message, however, gives clear indications of what is involved. The study of such cases will be taken up in its proper place. At the moment it is only necessary to point out that these are *code* messages and not *cipher*, and it is for this reason that in Pars. 12 and 13 the words "cipher" and "cipher messages" are used, the word "cryptogram" being used only where technically correct.

c. Secondly, there come the unusual and borderline cases, including cryptograms whose nature and type can *not* be ascertained from frequency distributions. Here, the cryptograms are technically not ciphers but special forms of disguised secret writings which are rarely susceptible of being classed as transposition or substitution. These include a large share of the cases wherein the cryptographic messages are disguised and carried under an external, innocuous text which is innocent and seemingly without cryptographic content—for instance, in a message wherein specific letters are indicated in a way not open to suspicion under censorship, these letters being intended to constitute the letters of the cryptographic message and the other letters constituting "dummies." Obviously, no amount of frequency tabulations will avail a competent, expert cryptanalyst in demonstrating or disclosing the presence of a cryptographic message, written and secreted within the "open" message, which serves but as an envelop and disguise for its authentic or real import. Certainly, such frequency tabulations can disclose the existence *neither* of substitution *nor* transposition in these cases, since both forms are absent. Another very popular method that resembles the method mentioned above has for its basis a simple grille. The whole words forming the secret text are inserted within perforations cut in the paper and the remaining space filled carefully, using "nulls" and "dummies", making a seemingly innocuous, ordinary message. There are other methods of this general type which can obviously neither be detected nor cryptanalyzed, using the principles of frequency of recurrences and repetition. These can not be further discussed herein, but at a subsequent date a special text may be written for their handling.[2]

[1] See Sec. XV, *Elementary Military Cryptography*.

[2] The subparagraph which the student has just read (47c) contains a hidden cryptographic message. With the hints given in Par. 35*e* let the student see if he can find it.

48. Ciphers employing characters other than letters or figures.—*a.* In view of the foregoing remarks, when so-called symbol ciphers, that is, ciphers employing peculiar symbols, signs of punctuation, diacritical marks, figures of "dancing men", and so on are encountered in practical work nowadays, they are almost certain to be simple, monoalphabetic ciphers. They are adequately described in romantic tales,[3] in popular books on cryptography, and in the more common types of magazine articles. No further space need be given ciphers of this type in this text, not only because of their simplicity but also because they are encountered in military cryptography only in sporadic instances, principally in censorship activities. Even in the latter cases, it is usually found that such ciphers are employed in "intimate" correspondence for the exchange of sentiments that appear less decorous when set forth in plain language. They are very seldom used by authentic enemy agents. When such a cipher is encountered nowadays it may practically always be regarded as the work of the veriest tyro, when it is not that of a "crank" or a mentally-deranged person.

b. The usual preliminary procedure in handling such cases, where the symbols may be somewhat confusing to the mind because of their unfamiliar appearance to the eye, is to substitute letters for them consistently throughout the message and then treat the resulting text as an ordinary cryptogram composed of letters is treated. This procedure also facilitates the construction of the necessary frequency distributions, which would be tedious to construct by using symbols.

c. A final word must be said on the subject of symbol ciphers by way of caution. When symbols are used to replace letters, syllables, and entire words, then the systems approach code methods in principle, and can become difficult of solution.[4] The logical extension of the use of symbols in such a form of writing is the employment of arbitrary characters for a specially developed "shorthand" system bearing little or no resemblance to well-known, and therefore nonsecret, systems of shorthand, such as Gregg, Pitman, etc. Unless a considerable amount of text is available for analysis, a privately-devised shorthand may be very difficult to solve. Fortunately, such systems are rarely encountered in military cryptography. They fall under the heading of cryptographic curiosities, of interest to the cryptanalyst in his leisure moments.[5]

d. In practical cryptography today, as has been stated above, the use of characters other than the 26 letters of the English alphabet is comparatively rare. It is true that there are a few governments which still adhere to systems yielding cryptograms in groups of figures. These are almost in every case code systems and will be treated in their proper place. In some cases cipher systems, or systems of enciphering code are used which are basically mathematical in character and operation, and therefore use numbers instead of letters. Some persons are inclined toward the use of numbers rather than letters because numbers lend themselves much more readily to certain arithmetical operations such as addition, subtraction, and so on, than do letters.[6] But there is usually added some final process whereby the figure groups are converted into letter groups, for the sake of economy in transmission.

---

[3] The most famous: Poe's *The Gold Bug;* Arthur Conan Doyle's *The Sign of Four.*

[4] The use of symbols for abbreviation and speed in writing goes back to the days of antiquity. Cicero is reported to have drawn up "a book like a dictionary, in which he placed before each word the notation (symbol) which should represent it, and so great was the number of notations and words that whatever could be written in Latin could be expressed in his notations."

[5] An example is found in the famous Pepys Diary, which was written in shorthand, purely for his own eyes by Samuel Pepys (1633–1703). "He wrote it in Shelton's system of tachygraphy (1641), which he complicated by using foreign languages or by varieties of his own invention whenever he had to record passages least fit to be seen by his servants, or by 'all the world.' "

[6] But, this of course, is because we are taught arithmetic by using numbers, based upon the decimal system as a rule. By special training one *could* learn to perform the usual "arithmetical" operations using letters. For example, using our English alphabet of 26 letters, where A=1, B=2, C=3, etc., it is obvious that A+B=C, just as 1+2=3; $(A+B)^2=I$, etc. This sort of cryptographic arithmetic could be learned by rote, just as multiplication tables are learned.

e. The only notable exceptions to the statement contained in the first sentence of the preceding subparagraph are those of Russian messages transmitted in the Russian Morse alphabet and Japanese messages transmitted in the Kata Kana Morse alphabet. As regards Chinese, which is not an alphabetical language and comprises some 40,000 ideographs, since the Morse telegraph code comprises only some 40 combinations, telegrams in Chinese are usually prepared by means of codes which permit of substituting arbitrarily-assigned code groups for the characters. Usually the code groups consist of figures. One such code known as the *Official Chinese Telegraph Code*, has about 10,000 4-figure groups, beginning with 0001, and these are arranged so that there are 100 characters on each page. Sometimes, for purposes of secrecy or economy, these figure groups are enciphered and converted in letter groups.

49. Concluding remarks concerning monoalphabetic substitution.—*a.* The alert student will have by this time gathered that the solution of monoalphabetic substitution ciphers of the simple or fixed type are particularly easy to solve, once the underlying principles are thoroughly understood. As in other arts, continued practice with examples leads to facility and skill in solution, especially where the student concentrates his attention upon traffic all of the same general nature, so that the type of text which he is continually encountering becomes familiar to him and its peculiarities or characteristics of construction give clues for short cuts to solution. It is true that a knowledge of the general phraseology of messages, the kind of words used, their sequences, and so on, is of very great assistance in practical work in all fields of cryptanalysis. The student is urged to note particularly these finer details in the course of his study.

b. Another thing which the student should be on the lookout for in simple monoalphabetic substitution is the consecutive use of several different mixed cipher alphabets in a single long message. Obviously, a single, composite frequency distribution for the whole message will not show the characteristic crest and trough appearance of a simple monoalphabetic cipher, since a given cipher letter will represent different plain-text letters in different parts of the message. But if the cryptanalyst will carefully observe the distribution *as it is being compiled*, he will note that at first it presents the characteristic crest and trough appearance of monoalphabeticity, and that after a time it begins to lose this appearance. If possible he should be on the lookout for some peculiarity of grouping of letters which serves as an indicator for the shift from one cipher alphabet to the next. If he finds such an indicator he should begin a second distribution from that point on, and proceed until another shift or indicator is encountered. By thus isolating the different portions of the text, and restricting the frequency distributions to the separate monoalphabets, the problem may be treated then as an ordinary simple monoalphabetic substitution. Consideration of these remarks in connection with instances of this kind leads to the comment that it is often more advisable for the cryptanalyst to compile his own data, than to have the latter prepared by clerks, especially when studying a system *de novo*. For observations which will certainly escape an untrained clerk can be most useful and may indeed facilitate solution. For example, in the case under consideration, if a clerk should merely hand the uniliteral distribution to the cryptanalyst, the latter might be led astray; the appearance of the composite distribution might convince him that the cryptogram is a good deal more complicated than it really is.

c. Monoalphabetic substitution with variants represents an extension of the basic principle, with the intention of masking the characteristic frequencies resulting from a strict monoalphabeticity, by means of which solutions are rather readily obtained. Some of the subterfuges applied on the establishment of variant or multiple values are simple and more or less fail to serve the purpose for which they are intended; others, on the contrary, may interpose serious difficulties to a straightforward solution. But in no case may the problem be considered of more than ordinary difficulty. Furthermore, it should be recognized that where these subterfuges

are really adequate to the purpose, the complications introduced are such that the practical manipulation of the system becomes as difficult for the cryptographer as for the cryptanalyst.

d. As already mentioned in monoalphabetic substitution with variants it is most common to employ figures or groups of figures. The reason for this is that the use of numerical groups seems more natural or easier to the uninitiated than does the use of varying combinations of letters. Moreover, it is easy to draw up cipher alphabets in which some of the letters are represented by single digits, others by pairs of digits. Thus, the decomposition of the cipher text which is an irregular intermixture of uniliteral and multiliteral equivalents, is made more complicated and correspondingly difficult for the cryptanalyst, who does not known which digits are to be used separately, which in pairs.

e. A few words may be added here in regard to a method which often suggests itself to laymen. This consists in using a book possessed by all the correspondents and indicating the letters of the message by means of numbers referring to specific letters in the book. One way consists in selecting a certain page and then giving the line number and position of the letter in the line, the page number being shown by a single initial indicator. Another way is to use the entire book, giving the cipher equivalents in groups of three numbers representing page, line, and number of letter. (Ex.: 75–8–10 means page 75, 8th line, 10th letter in the line.) Such systems are, however, extremely cumbersome to use and, when the cryptographing is done carelessly, can be solved. The basis for solution in such cases rests upon the use of adjacent letters on the same line, the accidental repetitions of certain letters, and the occurrence of unenciphered words in the messages, when laziness or fatigue intervenes in the cryptographing.[7]

f. It may also be indicated that human nature and the fallibility of cipher clerks is such that it is rather rare for an encipherer to make full use of the complement of variants placed at his disposal. The result is that in most cases certain of the equivalents will be used so much more often than others that diversities in frequencies will soon manifest themselves, affording important data for attack by the cryptanalyst.

g. In the World War the cases where monoalphabetic substitution ciphers were employed in actual operations on the Western Front were exceedingly rare because the majority of the belligerents had a fair knowledge of cryptography. On the Eastern Front, however, the extensive use, by the poorly prepared Russian Army, of monoalphabetic ciphers in the fall of 1914 was an important, if not the most important, factor in the success of the German operations during the Battle of Tannenberg.[8] It seems that a somewhat more secure cipher system was authorized, but proved too difficult for the untrained Russian cryptographic and radio personnel. Consequently, recourse was had to simple substitution ciphers, somewhat interspersed with plain text, and sometimes to messages completely in plain language. The damage which this faulty use of cryptography did to the Russian Army and thus to the Allied cause is incalculabl

h. Many of the messages found by censors in letters sent by mail during the World War were cases of monoalphabetic substitution, disguised in various ways.

[7] In 1915 the German Government conspired with a group of Hindu revolutionaries to stir up a rebellion in India, the purpose being to cause the withdrawal of British troops from the Western Front. Hindu conspirators in the United States were given money to purchase arms and ammunition and to transport them to India. For communication with their superiors in Berlin the conspirators used, among others, the system described in this paragraph. A 7-page typewritten letter, built up from page, line, and letter-number references to a book known only to the communicants, was intercepted by the British and turned over to the United States Government for use in connection with the prosecution of the Hindus for violating our neutrality. The author solved this message without the book in question, by taking full advantage of the clues referred to.

[8] Gyldén, Yves. *Chifferbydernas Insatser I Världskriget Till Lands*, Stockholm, 1931. A translation under the title *The Contribution of the Cryptographic Bureaus in the World War*, appeared in the Signal Corps Bulletin in seven successive installments, from November–December 1933 to November–December 1934, inclusive.

Nikolaieff, A. M. *Secret Causes of German success on the Eastern Front.* Coast Artillery Journal, September–October, 1935.

50. Analytical key for cryptanalysis.—*a.* It may be of assistance to indicate, by means of an outline, the relationships existing among the various cryptographic systems thus far considered. This graphic outline will be augmented from time to time as the different cipher systems are examined, and will constitute what has already been alluded to in Par. 6*d* and there termed an analytical key for cryptanalysis.[9] Fundamentally its nature is that of a schematic classification of the different systems examined. The analytical key forms an insert at the end of the book.

b. Note, in the analytical key, the rather clear-cut, dichotomous method of treatment; that is, classification by subdivision into pairs. For example, in the very first step there are only two alternatives: the cryptogram is either (1) cipher, or (2) code. If it is cipher, it is either (1) substitution, (2) transposition. If it is a substitution cipher, it is either (1) monographic, or (2) polygraphic—and so on. If the student will study the analytical key attentively, it will assist him in fixing in mind the manner in which the various systems covered thus far are related to one another, and this will be of benefit in clearing away some of the mental fog or haziness from which he is at first apt to suffer.

c. The numbers in parentheses refer to specific paragraphs in this text, so that the student may readily turn to the text for detailed information or for purposes of refreshing his memory as to procedure.

d. In addition to these reference numbers there have been affixed to the successive steps in the dichotomy, numbers that mark the "routes" on the cryptanalytic map (the analytical key) which the student cryptanalyst should follow if he wishes to facilitate his travels along the rather complicated and difficult road to success in cryptanalysis, in somewhat the same way in which an intelligent motorist follows the routes indicated on a geographical map if he wishes to facilitate his travels along unfamiliar roads. The analogy is only partially valid, however. The motorist usually knows in advance the distant point which he desires to reach and he proceeds thereto by the best and shortest route, which he finds by observing the route indications on a map and following the route markers on the road. Occasionally he encounters a detour but these are unexpected difficulties as a rule. Least of all does he anticipate any necessity for journeys down what may soon turn out to be blind alleys and "dead-end" streets, forcing him to double back on his way. Now the cryptanalyst also has a distant goal in mind—the solution of the cryptogram at hand—but he does not know at the outset of his journey the exact spot where it is located on the cryptanalytic map. The map contains many routes and he proceeds

[9] This analytical key is quite analogous to the analytical keys usually found in the handbooks biologists commonly employ in the classification and identification of living organisms. In fact, there are several points of resemblance between, for example, that branch of biology called taxonomic botany and cryptanalysis. In the former the first steps in the classificatory process are based upon observation of externally quite marked differences; as the process continues, the observational details become finer and finer, involving more and more difficulties as the work progresses. Towards the end of the work the botanical taxonomist may have to dissect the specimen and study internal characteristics. The whole process is largely a matter of painstaking, accurate observation of data and drawing proper conclusions therefrom. Except for the fact that the botanical taxonomist depends almost entirely upon ocular observation of characteristics while the cryptanalyst in addition to observation must use some statistics, the steps taken by the former are quite similar to those taken by the latter. It is only at the very end of the work that a significant dissimilarity between the two sciences arises. If the botanist makes a mistake in observation or deduction, he merely fails to identify the specimen correctly; he has an "answer"—but the answer is wrong. He may not be cognizant of the error; however, other more skillful botanists will find him out. But if the cryptanalyst makes a mistake in observation or deduction, he fails to get any "answer" at all; he needs nobody to tell him he has failed. Further, there is one additional important point of difference. The botanist is studying a bit of Nature—and she does not consciously interpose obstacles, pitfalls, and dissimulations in the path of those trying to solve her mysteries. The cryptanalyst, on the other hand, is studying a piece of writing prepared with the express purpose of preventing its being read by any persons for whom it is not intended. The obstacles, pitfalls, and dissimulations are here consciously interposed by the one who cryptographed the message. These, of course, are what make cryptanalysis different and difficult.

to test them one by one, in a successive chain. He encounters many blind alleys and dead-end streets, which force him to retrace his steps; he makes many detours and jumps many hurdles. Some of these retracings of steps, doubling back on his tracks, jumping of hurdles, and detours are unavoidable, but a few are avoidable. If properly employed, the analytical key will help the careful student to avoid those which should and can be avoided; if it does that much it will serve the principal purpose for which it is intended.

e. The analytical key may, however, serve another purpose of a somewhat different nature. When a multitude of cryptographic systems of diverse types must be filed in some systematic manner apart from the names of the correspondents or other reference data, or if in conducting instructional activities classificatory designations are desirable, the reference numbers on the analytical key may be made to serve as "type numbers." Thus, instead of stating that a given cryptogram is a keyword-systematically-mixed-uniliteral-monoalphabetic-monographic substitution cipher one may say that it is a "Type 901 cryptogram."

f. The method of assigning type numbers is quite simple. If the student will examine the numbers he will note that successive levels in the dichotomy are designated by successive hundreds. Thus, the first level, the classification into cipher and code is assigned the numbers 101 and 102. On the second level, under cipher, the classification into monographic and polygraphic systems is assigned the numbers 201 and 202, etc. Numbers in the same hundreds apply therefore to systems at the same level in the classification. There is no particular virtue in this scheme of assigning type numbers except that it provides for a considerable degree of expansion in future studies.

APPENDIX 1

(105)

APPENDIX

Table No.		Page
1–A.	Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters. Arranged alphabetically	108
1–B.	Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters. Arranged according to frequency	109
1–C.	Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters	110
2–A.	Absolute frequency of letters appearing in the combined five sets of messages totalling 50,000 letters. Arranged alphabetically	109
2–B.	Absolute frequency of letters appearing in the combined five sets of messages totalling 50,000 letters. Arranged according to frequency	110
2–C.	Absolute frequency of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in the combined five sets of messages totalling 50,000 letters	110
2–D.	Absolute frequencies of letters as initial letters of 10,000 words found in Government plain-text telegrams. (1) Arranged alphabetically, and (2) according to absolute frequencies	111
2–E.	Absolute frequencies of letters as final letters of 10,000 words found in Government plain-text telegrams. (1) Arranged alphabetically, and (2) according to absolute frequencies	111
3.	Relative frequencies of letters appearing in 1,000 letters based upon Table 2. (1) Arranged alphabetically, (2) according to absolute frequency, (3) vowels, (4) high frequency consonants, (5) medium frequency consonants, and (6) low frequency consonants	112
4.	Frequency distribution for 10,000 letters of literary English. (1) Arranged alphabetically, and (2) according to absolute frequencies	113
5.	Frequency distribution for 10,000 letters of telegraphic English. (1) Arranged alphabetically, and (2) according to absolute frequencies	113
6.	Frequency distribution of digraphs, based on 50,000 letters of Government plain-text telegrams, reduced to 5,000 digraphs	113
7–A.	The 438 different digraphs of Table 6. Arranged according to their absolute frequencies	114
7–B.	The 18 digraphs composing 25% of the digraphs in Table 6. Arranged alphabetically according to their initial letters, (1) and according to their final letters (2) and according to their absolute frequencies	116
7–C.	The 53 digraphs composing 50% of the digraphs in Table 6. Arranged alphabetically according to their initial letters, (1) and according to their final letters (2) and according to their absolute frequencies	117
7–D.	The 117 digraphs composing 75% of the digraphs in Table 6. Arranged alphabetically according to their initial letters, (1) and according to their final letters (2) and according to their absolute frequencies	118
7–E.	All the 438 digraphs of Table 6. Arranged first alphabetically according to their initial letters and then alphabetically according to their final letters	119
	(See Table 6. Read across the rows)	113
8.	The 438 different digraphs of Table 6. Arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter	120
9–A.	The 438 different digraphs of Table 6. Arranged first alphabetically according to their final letters and then according to their absolute frequencies	123
9–B.	The 18 digraphs composing 25% of the 5,000 digraphs of Table 6. Arranged alphabetically according to their final letters, (1) and according to their initial letters, (2) and according to their absolute frequencies	126
9–C.	The 53 digraphs composing 50% of the 5,000 digraphs of Table 6. Arranged alphabetically according to their final letters, (1) and according to their initial letters, (2) and according to their absolute frequencies	126
9–D.	The 117 digraphs composing 75% of the 5,000 digraphs of Table 6. Arranged alphabetically according to their final letters, (1) and according to their initial letters, (2) and according to their absolute frequencies	127
9–E.	All the 438 different digraphs of Table 6. Arranged alphabetically first according to their final letters and then according to their initial letters	129
	(See Table 6. Read down the columns)	113

Table No.		Page
10–	The 56 trigraphs appearing 100 or more times in the 50,000 letters of government plain-text telegrams—	
–A.	Arranged according to their absolute frequencies	129
–B.	Arranged first alphabetically according to their initial letters and then according to their absolute frequencies	130
–C.	Arranged first alphabetically according to their central letters and then according to their absolute frequencies	130
–D.	Arranged first alphabetically according to their final letters and then according to their absolute frequencies	131
11–	The 54 tetragraphs appearing 50 or more times in the 50,000 letters of government plain-text telegrams—	
–A.	Arranged according to their absolute frequencies	132
–B.	Arranged first alphabetically according to their initial letters and then according to their absolute frequencies	132
–C.	Arranged first alphabetically according to their second letters and then according to their absolute frequencies	133
–D.	Arranged first alphabetically according to their third letters and then according to their absolute frequencies	133
–E.	Arranged first alphabetically according to their final letters and then according to their absolute frequencies	134
12.	Average and mean lengths of words	135

TABLE 1–A.—*Absolute frequencies of letters appearing in five sets of Governmental plain-text telegrams, each set containing 10,000 letters, arranged alphabetically*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
A	738	A	783	A	681	A	740	A	741
B	104	B	103	B	98	B	83	B	99
C	319	C	300	C	288	C	326	C	301
D	387	D	413	D	423	D	451	D	448
E	1, 367	E	1, 294	E	1, 292	E	1, 270	E	1, 275
F	253	F	287	F	308	F	287	F	281
G	166	G	175	G	161	G	167	G	150
H	310	H	351	H	335	H	349	H	349
I	742	I	750	I	787	I	700	I	697
J	18	J	17	J	10	J	21	J	16
K	36	K	38	K	22	K	21	K	31
L	365	L	393	L	333	L	386	L	344
M	242	M	240	M	238	M	249	M	268
N	786	N	794	N	815	N	800	N	780
O	685	O	770	O	791	O	756	O	762
P	241	P	272	P	317	P	245	P	260
Q	40	Q	22	Q	45	Q	38	Q	30
R	760	R	745	R	762	R	735	R	786
S	658	S	583	S	585	S	628	S	604
T	936	T	879	T	894	T	958	T	928
U	270	U	233	U	312	U	247	U	238
V	163	V	173	V	142	V	133	V	155
W	166	W	163	W	136	W	133	W	182
X	43	X	50	X	44	X	53	X	41
Y	191	Y	155	Y	179	Y	213	Y	229
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10, 000		10, 000		10, 000		10, 000		10, 000

TABLE 2-A.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters, arranged alphabetically*

A	3,683	G	819	L	1,821	Q	175	V	766
B	487	H	1,694	M	1,237	R	3,788	W	780
C	1,534	I	3,676	N	3,975	S	3,058	X	231
D	2,122	J	82	O	3,764	T	4,595	Y	967
E	6,498	K	148	P	1,335	U	1,300	Z	49
F	1,416								

TABLE 1-B.—*Absolute frequencies of letters appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters, arranged according to frequency*

Set No. 1		Set No. 2		Set No. 3		Set No. 4		Set No. 5	
Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency	Letter	Absolute Frequency
E	1,367	E	1,294	E	1,292	E	1,270	E	1,275
T	936	T	879	T	894	T	958	T	928
N	786	N	794	N	815	N	800	R	786
R	760	A	783	O	791	O	756	N	780
I	742	O	770	I	787	A	740	O	762
A	738	I	750	R	762	R	735	A	741
O	685	R	745	A	681	I	700	I	697
S	658	S	583	S	585	S	628	S	604
D	387	D	413	D	423	D	451	D	448
L	365	L	393	H	335	L	386	H	349
C	319	H	351	L	333	H	349	L	344
H	310	C	300	P	317	C	326	C	301
U	270	F	287	U	312	F	287	F	281
F	253	P	272	F	308	M	249	M	268
M	242	M	240	C	288	U	247	P	260
P	241	U	233	M	238	P	245	U	238
Y	191	G	175	Y	179	Y	213	Y	229
G	166	V	173	G	161	G	167	W	182
W	166	W	163	V	142	V	133	V	155
V	163	Y	155	W	136	W	133	G	150
B	104	B	103	B	98	B	83	B	99
X	43	X	50	Q	45	X	53	X	41
Q	40	K	38	X	44	Q	38	K	31
K	36	Q	22	K	22	K	21	Q	30
J	18	J	17	J	10	J	21	J	16
Z	14	Z	17	Z	2	Z	11	Z	5
Total	10,000		10,000		10,000		10,000		10,000

148274—38——8

TABLE 1–C.—*Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in five sets of Government plain-text telegrams, each set containing 10,000 letters*

Set No.	Vowels	High Frequency Consonants	Medium Frequency Consonants	Low Frequency Consonants
1	3,993	3,527	2,329	151
2	3,985	3,414	2,457	144
3	4,042	3,479	2,356	123
4	3,926	3,572	2,358	144
5	3,942	3,546	2,389	123
Total [1]	19,888	17,538	11,889	685

[1] Grand total, 50,000.

TABLE 2–B.—*Absolute frequencies of letters appearing in the combined five sets of messages totalling 50,000 letters arranged according to frequencies*

E	6,498	I	3,676	C	1,534	Y	967	X	231
T	4,595	S	3,058	F	1,416	G	819	Q	175
N	3,975	D	2,122	P	1,335	W	780	K	148
R	3,788	L	1,821	U	1,300	V	766	J	82
O	3,764	H	1,694	M	1,237	B	487	Z	49
A	3,683								

TABLE 2–C.—*Absolute frequencies of vowels, high frequency consonants, medium frequency consonants, and low frequency consonants appearing in the combined five sets of messages totalling 50,000 letters*

Vowels	19,888
High Frequency Consonants (D, N, R, S, and T)	17,538
Medium Frequency Consonants (B, C, F, G, H, L, M, P, V, and W)	11,889
Low Frequency Consonants (J, K, Q, X, and Z)	685
Total	50,000

TABLE 2–D.—*Absolute frequencies of letters as initial letters of 10,000 words found in Government plain-text telegrams*

(1) ARRANGED ALPHABETICALLY

A	905	G	109	L	196	Q	30	V	77
B	287	H	272	M	384	R	611	W	320
C	664	I	344	N	441	S	965	X	4
D	525	J	44	O	646	T	1,253	Y	88
E	390	K	23	P	433	U	122	Z	12
F	855								
								Total	10,000

(2) ARRANGED ACCORDING TO ABSOLUTE FREQUENCIES

T	1,253	R	611	M	384	L	196	J	44
S	965	D	525	I	344	U	122	Q	30
A	905	N	441	W	320	G	109	K	23
F	855	P	433	B	287	Y	88	Z	12
C	664	E	390	H	272	V	77	X	4
O	646								
								Total	10,000

TABLE 2–E.—*Absolute frequencies of letters as final letters of 10,000 words found in Government plain-text telegrams*

(1) ARRANGED ALPHABETICALLY

A	269	G	225	L	354	Q	8	V	4
B	22	H	450	M	154	R	769	W	45
C	86	I	22	N	872	S	962	X	116
D	1,002	J	6	O	575	T	1,007	Y	866
E	1,628	K	53	P	213	U	31	Z	9
F	252								
								Total	10,000

(2) ARRANGED ACCORDING TO ABSOLUTE FREQUENCIES

E	1,628	R	769	F	252	C	86	I	22
T	1,007	O	575	G	225	K	53	Z	9
D	1,002	H	450	P	213	W	45	Q	8
S	962	L	354	M	154	U	31	J	6
N	872	A	269	X	116	B	22	V	4
Y	866								
								Total	10,000

TABLE 3.—*Relative frequencies of letters appearing in 1,000 letters based upon Table 2-B*

(1) ARRANGED ALPHABETICALLY

A	73.66	G	16.38	L	36.42	Q	3.50	V	15.32
B	9.74	H	33.88	M	24.74	R	75.76	W	15.60
C	30.68	I	73.52	N	79.50	S	61.16	X	4.62
D	42.44	J	1.64	O	75.28	T	91.90	Y	19.34
E	129.96	K	2.96	P	26.70	U	26.00	Z	.98
F	28.32								
								Total	1,000.00

(2) ARRANGED ACCORDING TO FREQUENCY

E	129.96	I	73.52	C	30.68	Y	19.34	X	4.62
T	91.90	S	61.16	F	28.32	G	16.38	Q	3.50
N	79.50	D	42.44	P	26.70	W	15.60	K	2.96
R	75.76	L	36.42	U	26.00	V	15.32	J	1.64
O	75.28	H	33.88	M	24.74	B	9.74	Z	.98
A	73.66								
								Total	1,000.00

(3) VOWELS

A	73.66
E	129.96
I	73.52
O	75.28
U	26.00
Y	19.34
Total	397.76

(4) HIGH-FREQUENCY CONSONANTS

D	42.44
N	79.50
R	75.76
S	61.16
T	91.90
Total	350.76

(5) MEDIUM-FREQUENCY CONSONANTS

B	9.74
C	30.68
F	28.32
G	16.38
H	33.88
L	36.42
M	24.74
P	26.70
V	15.32
W	15.60
Total	237.78

(6) LOW-FREQUENCY CONSONANTS

X	4.62
Q	3.50
K	2.96
J	1.64
Z	.98
Total	13.70
Total (3), (4), (5), (6)	1,000.00

First Letter																															
B	4				18				2	1		6	1		4			2	1	1	2				7		49	14
C	20		3	1	32	1		14	7		4	5	1	1	41			4	1	14	4		1		1		155	8
D	32	4	4	8	33	8	2	2	27	1		3	5	4	16	5	2	12	13	15	5	3	4		1		209	3
E	35	4	32	60	42	18	4	7	27	1		29	14	111	12	20	12	87	54	37	3	20	7	7	4	1	648	1
F	5		2	1	10	11	1		39			2	1		40	1		9	3	11	3		1		1		141	9
G	7		2	1	14	2	1	20	5	1		2	1	3	6	2		5	3	4	2		1				82	7
H	20	1	3	2	20	5			33			1	2	3	20	1	1	17	4	28	8		1		1		171	7
I	8	2	22	6	13	10	19				2	23	9	75	41	7		27	35	27		25		15		2	368	7
J	1				2										2						2						7	22
K	1		1		6				2			1		1					1								13	19
L	28	3	3	9	37	3	1	1	20			27	2	1	13	3		2	6	8	2	2	2		10		183	5
M	36	6	3	1	26	1		1	9				13		10	8		2	4	2	2				2		126	10
N	26	2	19	52	57	9	27	4	30	1	2	5	5	8	18	3	1	4	24	82	7	3	3		5		397	2
O	7	4	8	12	3	25	2	3	5	1	2	19	25	77	6	25		64	14	19	37	7	8	1	2		376	2
P	14	1	1	1	23	2		3	6			13	4	1	17	11		18	6	8	3	1	1		1		135	6
Q													1					1			15						17	23
R	39	2	9	17	98	6	7	3	30	1	1	5	9	7	28	13		11	31	42	5	5	4		9		382	3
S	24	3	13	5	49	12	2	26	34		1	2	3	4	15	10		5	19	63	11	1	4		1		307	4
T	28	3	6	6	71	7	1	78	45			5	6	7	50	2	1	17	19	19	5		36		41	1	454	4
U	5	3	3	3	11	1	8		5			6	5	21	1	2		31	12	12		1					130	9
V	6				57				12						1					1							77	21
W	12				22			4	13			1		2	19			1	1						1		76	16
X	2		2	1	1	1		1	2					1	1	2		1	1	7							23	13
Y	6	2	4	4	9	11	1	1	3			2	2	6	10	3		4	11	15	1		1				96	7
Z	1				2				1																		4	23
Total	370	46	154	217	657	137	82	170	374	8	14	189	123	397	373	130	17	368	304	462	130	75	77	23	99	4	5,000	
Blanks	1	11	6	7	1	7	12	10	3	18	19	6	6	7	3	8	21	4	4	5	7	15	11	23	10	23		248

148274—38 (Face p. 113)

TABLE 4.—*Frequency distribution for 10,000 letters of literary English, as compiled by Hitt*[1]

(1) ALPHABETICALLY ARRANGED

A	778	G	174	L	372	Q	8	V	112
B	141	H	595	M	288	R	651	W	176
C	296	I	667	N	686	S	622	X	27
D	402	J	51	O	807	T	855	Y	196
E	1,277	K	74	P	223	U	308	Z	17
F	197								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,277	R	651	U	308	Y	196	K	74
T	855	S	622	C	296	W	176	J	51
O	807	H	595	M	288	G	174	X	27
A	778	D	402	P	223	B	141	Z	17
N	686	L	372	F	197	V	112	Q	8
I	667								

TABLE 5.—*Frequency distribution for 10,000 letters of telegraphic English as compiled by Hitt*

(1) ALPHABETICALLY ARRANGED

A	813	G	201	L	392	Q	38	V	136
B	149	H	386	M	273	R	677	W	166
C	306	I	711	N	718	S	656	X	51
D	417	J	42	O	844	T	634	Y	208
E	1,319	K	88	P	243	U	321	Z	6
F	205								

(2) ARRANGED ACCORDING TO FREQUENCY

E	1,319	S	656	U	321	F	205	K	88
O	844	T	634	C	306	G	201	X	51
A	813	D	417	M	273	W	166	J	42
N	718	L	392	P	243	B	149	Q	38
I	711	H	386	Y	208	V	136	Z	6
R	677								

[1] Hitt, Capt. Parker. *Manual for the Solution of Military Ciphers.* Army Service Schools Press, Fort Leavenworth, Kansas, 1916.

TABLE 7-A.—*The 438 different digraphs of table 6 arranged according to their absolute frequencies*

EN	111	EC	32	OL	19	US	12
RE	98	RS	31	OT	19	UT	12
ER	87	UR	31	TS	19	VI	12
NT	82	NI	30	WO	19	WA	12
TH	78	RI	30	BE	18	FF	11
ON	77	EL	29	EF	18	PP	11
IN	75	HT	28	NO	18	RR	11
TE	71	LA	28	PR	18	UE	11
AN	64	RO	28	AI	17	FT	11
OR	64	TA	28	HR	17	SU	11
ST	63			PO	17	YF	11
ED	60		[2] 2,495	RD	17	YS	11
NE	57	LL	27	TR	17	YO	10
VE	57	AD	27	DO	16	FE	10
ES	54	DI	27	DT	15	IF	10
ND	52	EI	27	IX	15	LY	10
TO	50	IR	27	QU	15	MO	10
SE	49	IT	27	SO	15	SP	10
		NG	27	YT	15	YE	9
	[1] 1,249	ME	26	AC	14	FR	9
AT	47	NA	26	AM	14	IM	9
TI	45	SH	26	CH	14	LD	9
AR	44	IV	25	CT	14	MI	9
EE	42	OF	25	EM	14	NF	9
RT	42	OM	25	GE	14	RC	9
AS	41	OP	25	OS	14	RM	9
CO	41	NS	24	PA	14	RY	9
IO	41	SA	24	PL	13	DD	8
TY	41	IL	23	RP	13	NN	8
FO	40	PE	23	SC	13	DF	8
FI	39	IC	22	WI	13	IA	8
RA	39	WE	22	MM	13	HU	8
ET	37	UN	21	DS	13	LT	8
OU	37	CA	20	AU	13	MP	8
LE	37	EP	20	IE	13	OC	8
MA	36	EV	20	LO	13	OW	8
TW	36	GH	20		[3] 3,745	PT	8
EA	35	HA	20			UG	8
IS	35	HE	20	AP	12	AV	7
SI	34	HO	20	DR	12	BY	7
DE	33	LI	20	EQ	12	CI	7
HI	33	SS	19	AY	12	EH	7
AL	32	TT	19	EO	12	OA	7
CE	32	IG	19	OD	12	EW	7
DA	32	NC	19	SF	12	EX	7

[1] The 18 digraphs above this line compose 25% of the total.
[2] The 53 digraphs above this line compose 50% of the total.
[3] The 117 digraphs above this line compose 75% of the total.

TABLE 7-A.—*The 438 different digraphs of table 6 arranged according to their absolute frequencies*—Continued

GA	7	SD	5	DV	3	KI	2
IP	7	SR	5	AA	3	LM	2
NU	7	TL	5	EU	3	LR	2
OV	7	TU	5	OE	3	LU	2
RG	7	UM	5	YI	3	LV	2
RN	7	AF	4	FS	3	LW	2
TE	7	BA	4	FU	3	MR	2
TN	7	BO	4	GN	3	MT	2
XT	7	CK	4	GS	3	MU	2
AB	6	CR	4	HC	3	MY	2
AG	6	CU	4	HN	3	NB	2
BL	6	DB	4	LB	3	NK	2
OO	6	DC	4	LC	3	OG	2
YA	6	DN	4	LF	3	OK	2
GO	6	DW	4	LP	3	PF	2
ID	6	EB	4	MC	3	RB	2
KE	6	EG	4	NP	3	SG	2
LS	6	EY	4	NV	3	SL	2
MB	6	GT	4	NW	3	TP	2
PI	6	HS	4	OH	3	UP	2
PS	6	MS	4	AH	2	WN	2
RF	6	NH	4	AK	2	XA	2
TC	6	NR	4	BI	2	XC	2
TD	6	OB	4	BR	2	XI	2
TM	6	PM	4	BU	2	XP	2
UL	6	RW	4	DG	2	YB	2
VA	6	SN	4	DH	2	YL	2
YN	6	SW	4	DO	2	YM	2
CL	5	WH	4	AO	2	ZE	2
DM	5	YC	4	OY	2	GG	1
DP	5	YD	4	FC	2	AJ	1
DU	5	YR	4	FL	2	BJ	1
OI	5	PH	3	GC	2	BM	1
UA	5	PU	3	GF	2	BS	1
UI	5	RH	3	GL	2	BT	1
FA	5	SB	3	GP	2	CD	1
GI	5	SM	3	GU	2	CF	1
GR	5	TB	3	HD	2	CM	1
HF	5	UB	3	HM	2	CN	1
NL	5	UC	3	IB	2	CS	1
NM	5	UD	3	IK	2	CW	1
NY	5	YP	3	IZ	2	CY	1
RL	5	CC	3	JE	2	DJ	1
RU	5	AW	3	JO	2	DY	1
RV	5	DL	3	JU	2	EJ	1

TABLE 7-A.—*The 138 different digraphs of table 6 arranged according to their absolute frequencies*—Continued

AE	1	HY	1	PD	1	WL	1
UO	1	JA	1	PN	1	WR	1
YU	1	KA	1	PV	1	WS	1
EZ	1	KC	1	PW	1	WY	1
FD	1	KL	1	PY	1	XD	1
FG	1	KN	1	QM	1	XE	1
FM	1	KS	1	QR	1	XF	1
FP	1	LG	1	RJ	1	XH	1
FW	1	LH	1	RK	1	XN	1
FY	1	LN	1	SK	1	XO	1
GD	1	MD	1	SV	1	XR	1
GJ	1	MF	1	SY	1	XS	1
GM	1	MH	1	TG	1	YG	1
GW	1	NJ	1	TQ	1	YH	1
HB	1	NQ	1	TZ	1	YW	1
HL	1	OJ	1	UF	1	ZA	1
HP	1	OX	1	UV	1	ZI	1
HQ	1	PB	1	VO	1		
HW	1	PC	1	VT	1	Total	5,000

TABLE 7-B.—*The 18 digraphs composing 25% of the digraphs in Table 6 arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS				(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES			
AN	64	ON	77	AN	64	ON	77
		OR	64			OR	64
ED	60	RE	98	EN	111	RE	98
EN	111			ER	87		
ER	87	SE	49	ED	60	SE	49
ES	54	ST	63	ES	54	ST	63
		TE	71			TH	78
IN	75	TH	78	IN	75	TE	71
		TO	50			TO	50
ND	52	VE	57	NT	82	VE	57
NE	57			NE	57		
NT	82	Total	1,249	ND	52	Total	1,249

TABLE 7–C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AL	32	MA	36
AN	64		
AR	44	ND	52
AS	41	NE	57
AT	47	NI	30
		NT	82
CE	32		
CO	41	ON	77
		OR	64
DA	32	OU	37
DE	33		
		RA	39
EA	35	RE	98
EC	32	RI	30
ED	60	RO	28
EE	42	RS	31
EL	29	RT	42
EN	111		
ER	87	SE	49
ES	54	SI	34
ET	37	ST	63
FI	39	TA	28
FO	40	TE	71
		TH	78
HI	33	TI	45
HT	28	TO	50
		TW	36
IN	75	TY	41
IO	41		
IS	35	UR	31
		VE	57
LA	28		
LE	37	Total	2,495

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64	MA	36
AT	47		
AR	44	NT	82
AS	41	NE	57
AL	32	ND	52
		NI	30
CO	41		
CE	32	ON	77
		OR	64
DE	33	OU	37
DA	32		
		RE	98
EN	111	RT	42
ER	87	RA	39
ED	60	RS	31
ES	54	RI	30
EE	42	RO	28
ET	37		
EA	35	ST	63
EC	32	SE	49
EL	29	SI	34
		TH	78
FO	40	TE	71
FI	39	TO	50
		TI	45
HI	33	TY	41
HT	28	TW	36
		TA	28
IN	75		
IO	41	UR	31
IS	35		
		VE	57
LE	37		
LA	28	Total	2,495

TABLE 7–D.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters——*

(1) AND ACCORDING TO THEIR FINAL LETTERS

AC	14	EP	20	LO	13	RI	30
AD	27	ER	87			RO	28
AI	17	ES	54	MA	36	RS	31
AL	32	ET	37	ME	26	RT	42
AM	14	EV	20				
AN	64			NA	26	SA	24
AR	44	FI	39	NC	19	SE	49
AS	41	FO	40	ND	52	SH	26
AT	47			NE	57	SI	34
AU	13	GE	14	NG	27	SO	15
		GH	20	NI	30	SS	19
BE	18			NO	18	ST	63
		HA	20	NS	24		
CA	20	HE	20	NT	82	TA	28
CE	32	HI	33			TE	71
CH	14	HO	20	OF	25	TH	78
CO	41	HR	17	OL	19	TI	45
CT	14	HT	28	OM	25	TO	50
				ON	77	TR	17
DA	32	IC	22	OP	25	TS	19
DE	33	IE	13	OR	64	TT	19
DI	27	IG	19	OS	14	TW	36
DO	16	IL	23	OT	19	TY	41
DS	13	IN	75	OU	37		
DT	15	IO	41			UN	21
		IR	27	PA	14	UR	31
EA	35	IS	35	PE	23		
EC	32	IT	27	PO	17	VE	57
ED	60	IV	25	PR	18		
EE	42	IX	15			WE	22
EF	18			QU	15	WO	19
EI	27	LA	28				
EL	29	LE	37	RA	39	YT	15
EM	14	LI	20	RD	17		
EN	111	LL	27	RE	98	Total	3,745

TABLE 7-D, Concluded.—*The 117 digraphs comprising 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their initial letters——*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

AN	64	EI	27	MA	36	RI	30
AT	47	EP	20	ME	26	RO	28
AR	44	EV	20			RD	17
AS	41	EF	18	NT	82		
AL	32	EM	14	NE	57	ST	63
AD	27			ND	52	SE	49
AI	17	FO	40	NI	30	SI	34
AC	14	FI	39	NG	27	SH	26
AM	14			NA	26	SA	24
AU	13	GH	20	NS	24	SS	19
		GE	14	NC	19	SO	15
BE	18			NO	18		
		HI	33			TH	78
		HT	28	ON	77	TE	71
CO	41	HA	20	OR	64	TO	50
CE	32	HE	20	OU	37	TI	45
CA	20	HO	20	OF	25	TY	41
CH	14	HR	17	OM	25	TW	36
CT	14			OP	25	TA	28
		IN	75	OL	19	TS	19
DE	33	IO	41	OT	19	TT	19
DA	32	IS	35	OS	14	TR	17
DI	27	IR	27				
DO	16	IT	27	PE	23	UR	31
DT	15	IV	25	PR	18	UN	21
DS	13	IL	23	PO	17		
		IC	22	PA	14	VE	57
EN	111	IG	19				
ER	87	IX	15	QU	15	WE	22
ED	60	IE	13			WO	19
ES	54			RE	98		
EE	42	LE	37	RT	42	YT	15
ET	37	LA	28	RA	39		
EA	35	LL	27	RS	31	Total	3,745
EC	32	LI	20				
EL	29	LO	13				

TABLE 7-E.—*All the 438 digraphs of Table 6, arranged first alphabetically according to their initial letters and then alphabetically according to their final letters.*

(SEE TABLE 6.—READ ACROSS THE ROWS)

TABLE 8.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

AN	64	CT	14	ED	60	GH	20
AT	47	CI	7	ES	54	GE	14
AR	44	CL	5	EE	42	GA	7
AS	41	CK	4	ET	37	GO	6
AL	32	CR	4	EA	35	GI	5
AD	27	CU	4	EC	32	GR	5
AI	17	CC	3	EL	29	GT	4
AC	14	CD	1	EI	27	GN	3
AM	14	CF	1	EP	20	GS	3
AU	13	CM	1	EV	20	GC	2
AP	12	CN	1	EF	18	GF	2
AY	12	CS	1	EM	14	GL	2
AV	7	CW	1	EO	12	GP	2
AB	6	CY	1	EQ	12	GU	2
AG	6			EH	7	GD	1
AF	4	DE	33	EW	7	GG	1
AA	3	DA	32	EX	7	GJ	1
AW	3	DI	27	EB	4	GM	1
AH	2	DO	16	EG	4	GW	1
AK	2	DT	15	EY	4		
AO	2	DS	13	EU	3		
AE	1	DR	12	EJ	1		
AJ	1	DD	8	EZ	1	HI	33
		DF	8			HT	28
BE	18	DM	5	FO	40	HA	20
BY	7	DP	5	FI	39	HE	20
BL	6	DU	5	FF	11	HO	20
BA	4	DB	4	FT	11	HR	17
BO	4	DC	4	FE	10	HU	8
BI	2	DN	4	FR	9	HF	5
BR	2	DW	4	FA	5	HS	4
BU	2	DL	3	FS	3	HC	3
BJ	1	DV	3	FU	3	HN	3
BM	1	DG	2	FC	2	HD	2
BS	1	DH	2	FL	2	HM	2
BT	1	DQ	2	FD	1	HB	1
		DJ	1	FG	1	HL	1
CO	41	DY	1	FM	1	HP	1
CE	32			FP	1	HQ	1
CA	20	EN	111	FW	1	HW	1
CH	14	ER	87	FY	1	HY	1

[1] For arrangement alphabetically first under intial letters and then under final letters, see Table 6.

TABLE 8, Contd.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

IN	75	LI	20	NE	57	OA	7
IO	41	LO	13	ND	52	OV	7
IS	35	LY	10	NI	30	OO	6
IR	27	LD	9	NG	27	OI	5
IT	27	LT	8	NA	26	OB	4
IV	25	LS	6	NS	24	OE	3
IL	23	LB	3	NC	19	OH	3
IC	22	LC	3	NO	18	OG	2
IG	19	LF	3	NF	9	OK	2
IX	15	LP	3	NN	8	OY	2
IE	13	LM	2	NU	7	OJ	1
IF	10	LR	2	NL	5	OX	1
IM	9	LU	2	NM	5		
IA	8	LV	2	NY	5	PE	23
IP	7	LW	2	NH	4	PR	18
ID	6	LG	1	NR	4	PO	17
		LH	1	NP	3	PA	14
IB	2	LN	1	NV	3	PL	13
IK	2			NW	3	PP	11
IZ	2	MA	36	NB	2	PT	8
		ME	26	NK	2	PI	6
JE	2	MM	13	NJ	1	PS	6
JO	2	MO	10	NQ	1	PM	4
JU	2	MI	9			PH	3
JA	1	MP	8	ON	77	PU	3
		MB	6	OR	64	PF	2
KE	6	MS	4	OU	37	PB	1
KI	2	MC	3	OF	25	PC	1
KA	1	MR	2	OM	25	PD	1
KC	1	MT	2	OP	25	PN	1
KL	1	MU	2	OL	19	PV	1
KN	1	MY	2	OT	19	PW	1
KS	1	MD	1	OS	14	PY	1
		MF	1				
LE	37	MH	1	OD	12	QU	15
LA	28			OC	8	QM	1
LL	27	NT	82	OW	8	QR	1

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6.

TABLE 8, Concluded.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies under each initial letter* [1]

RE	98	SR	5	US	12	XI	2
RT	42	SN	4	UT	12	XP	2
RA	39	SW	4	UE	11	XD	1
RS	31	SB	3	UG	8	XE	1
RI	30	SM	3	UL	6	XF	1
RO	28	SG	2	UA	5	XH	1
RD	17	SL	2	UI	5	XN	1
RP	13	SK	1	UM	5	XO	1
RR	11	SV	1	UB	3	XR	1
RC	9	SY	1	UC	3	XS	1
RM	9			UD	3		
RY	9	TH	78	UP	2	YT	15
RG	7	TE	71	UF	1	YF	11
RN	7	TO	50	UO	1	YS	11
RF	6	TI	45	UV	1	YO	10
RL	5	TY	41			YE	9
RU	5	TW	36	VE	57	YA	6
RV	5	TA	28	VI	12	YN	6
RW	4	TS	19	VA	6	YC	4
RH	3	TT	19	VO	1	YD	4
RB	2	TR	17	VT	1	YR	4
RJ	1	TF	7			YI	3
RK	1	TN	7	WE	22	YP	3
		TC	6	WO	19	YB	2
ST	63	TD	6	WI	13	YL	2
SE	49	TM	6	WA	12	YM	2
SI	34	TL	5	WH	4	YG	1
SH	26	TU	5	WN	2	YH	1
SA	24	TB	3	WL	1	YU	1
SS	19	TP	2	WR	1	YW	1
SO	15	TG	1	WS	1		
SC	13	TQ	1	WY	1	ZE	2
SF	12	TZ	1			ZA	1
SU	11			XT	7	ZI	1
SP	10	UR	31	XA	2		
SD	5	UN	21	XC	2	Total	5,000

[1] For arrangement alphabetically first under initial letters and then under final letters, see Table 6.

TABLE 9–A.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

RA	39	EC	32	RE	98	GF	2
MA	36	IC	22	TE	71	PF	1
EA	35	NC	19	NE	57	CF	2
DA	32	AC	14	VE	57	MF	1
LA	28	SC	13	SE	49	UF	1
TA	28	RC	9	EE	42	XF	1
NA	26	OC	8	LE	37		
SA	24	TC	6	DE	33		
CA	20	DC	4	CE	32	NG	27
HA	20	YC	4	ME	26	IG	19
PA	14	CC	3	PE	23	UG	8
WA	12	HC	3	WE	22	RG	7
IA	8	LC	3	HE	20	AG	6
GA	7	MC	3	BE	18	EG	4
OA	7	UC	3	GE	14	DG	2
VA	6	FC	2	IE	13	OG	2
YA	6	GC	2	UE	11	SG	2
FA	5	XC	2	FE	10	FG	1
UA	5	KC	1	YE	9	GG	1
BA	4	PC	1	KE	6	LG	1
AA	3			OE	3	TG	1
XA	2			JE	2	YG	1
JA	1	ED	60	ZE	2		
KA	1	ND	52	AE	1		
ZA	1	AD	27	XE	1		
		RD	17			TH	78
AB	6	OD	12			SH	26
MB	6	LD	9			GH	20
DB	4	DD	8	OF	25	CH	14
EB	4	ID	6	EF	18	EH	7
OB	4	TD	6	SF	12	NH	4
LB	3	SD	5	FF	11	WH	4
SB	3	YD	4	YF	11	OH	3
TB	3	UD	3	IF	10	PH	3
UB	3	HD	2	NF	9	RH	3
IB	2	CD	1	DF	8	AH	2
NB	2	FD	1	TF	7	DH	2
RB	2	GD	1	RF	6	LH	1
YB	2	MD	1	HF	5	MH	1
HB	1	PD	1	AF	4	XH	1
PB	1	XD	1	LF	3	YH	1

TABLE 9–A, Contd.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

TI	45	LL	27	AN	64	RP	13
FI	39	IL	23	UN	21	AP	12
SI	34	OL	19	NN	8	PP	11
HI	33	PL	13	RN	7	SP	10
NI	30	BL	6	TN	7	MP	8
RI	30	UL	6	YN	6	IP	7
DI	27	CL	5	DN	4	DP	5
EI	27	NL	5	SN	4	LP	3
LI	20	RL	5	GN	3	NP	3
AI	17	TL	5	HN	3	YP	3
WI	13	DL	3	WN	2	GP	2
VI	12	FL	2	CN	1	TP	2
MI	9	GL	2	KN	1	UP	2
CI	7	SL	2	LN	1	XP	2
PI	6	YL	2	PN	1	FP	1
GI	5	HL	1	XN	1	HP	1
OI	5	KL	1			EQ	12
UI	5	WL	1	TO	50	DQ	2
YI	3			CO	41	HQ	1
BI	2	OM	25	IO	41	NQ	1
KI	2	AM	14	FO	40	TQ	1
XI	2	EM	14	RO	28	ER	87
ZI	1	MM	13	HO	20	OR	64
		IM	9	WO	19	AR	44
AJ	1	RM	9	NO	18	UR	31
BJ	1	TM	6	PO	17	IR	27
DJ	1	DM	5	DO	16	PR	18
EJ	1	NM	5	SO	15	HR	17
GJ	1	UM	5	LO	13	TR	17
NJ	1	PM	4	EO	12	DR	12
OJ	1	SM	3	MO	10	RR	11
RJ	1	HM	2	YO	10	FR	9
		LM	2	GO	6	GR	5
CK	4	YM	2	OO	6	SR	5
AK	2	BM	1	BO	4	CR	4
IK	2	CM	1	AO	2	NR	4
NK	2	FM	1	JO	2	YR	4
OK	2	GM	1	UO	1	BR	2
RK	1	QM	1	VO	1	LR	2
SK	1			XO	1	MR	2
		EN	111			QR	1
AL	32	ON	77	OP	25	WR	1
EL	29	IN	75	EP	20	XR	1

TABLE 9-A, Concluded.—*The 438 different digraphs of Table 6, arranged first alphabetically according to their final letters, and then according to their absolute frequencies*

ES	54	OT	19	JU	2	PW	1
AS	41	TT	19	LU	2	YW	1
IS	35	DT	15	MU	2		
RS	31	YT	15	YU	1	IX	15
NS	24	CT	14			EX	7
SS	19	UT	12	IV	25	OX	1
TS	19	FT	11	EV	20		
OS	14	LT	8	AV	7	TY	41
DS	13	PT	8	OV	7	AY	12
US	12	XT	7	RV	5	LY	10
YS	11	GT	4	DV	3	RY	9
LS	6	MT	2	NV	3	BY	7
PS	6	BT	1	LV	2	NY	5
HS	4	VT	1	PV	1	EY	4
MS	4			SV	1	MY	2
FS	3	OU	37	UV	1	OY	2
GS	3	QU	15			CY	1
BS	1	AU	13	TW	36	DY	1
CS	1	SU	11	OW	8	FY	1
KS	1	HU	8	EW	7	HY	1
WS	1	NU	7	DW	4	PY	1
XS	1	DU	5	RW	4	SY	1
		RU	5	SW	4	WY	1
NT	82	TU	5	AW	3		
ST	68	CU	4	NW	3	IZ	2
AT	47	EU	3	LW	2	EZ	1
RT	42	FU	3	CW	1	TZ	1
ET	37	PU	3	FW	1		
HT	28	BU	2	GW	1	Total	5,000
IT	27	GU	2	HW	1		

TABLE 9-B.—*The 18 digraphs composing 25% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS

ED	60	IN	75
ND	52	ON	77
NE	57	TO	50
RE	98	ER	87
SE	49	OR	64
TE	71	ES	54
VE	57	NT	82
TH	78	ST	63
AN	64	Total	1,249
EN	111		

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

ED	60	IN	75
ND	52	AN	64
RE	98	TO	50
TE	71	ER	87
NE	57	OR	64
VE	57	ES	54
SE	49	NT	82
TH	78	ST	63
EN	111	Total	1,249
ON	77		

TABLE 9-C.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS

DA	32	RE	98	EN	111	IS	35
EA	35	SE	49	IN	75	RS	31
LA	28	TE	71	ON	77	AT	47
MA	36	VE	57	CO	41	ET	37
RA	39	TH	78	FO	40	HT	28
TA	28	FI	39	IO	41	NT	82
EC	32	HI	33	RO	28	RT	42
ED	60	NI	30	TO	50	ST	63
ND	52	RI	30	AR	44	OU	37
CE	32	SI	34	ER	87	TW	36
DE	33	TI	45	OR	64	TY	41
EE	42	AL	32	UR	31	Total	2,495
LE	37	EL	29	AS	41		
NE	57	AN	64	ES	54		

TABLE 9–C, Concluded.—*The 53 digraphs composing 50% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

RA	39	LE	37	ON	77	IS	35
MA	36	DE	33	IN	75	RS	31
EA	35	CE	32	AN	64		
DA	32					NT	82
LA	28	TH	78	TO	50	ST	63
TA	28			CO	41	AT	47
		TI	45	IO	41	RT	42
EC	32	FI	39	FO	40	ET	37
		SI	34	RO	28	HT	28
ED	60	HI	33				
ND	52	NI	30	ER	87	OU	37
		RI	30	OR	64		
RE	98			AR	44	TW	36
TE	71			UR	31		
NE	57	AL	32			TY	41
VE	57	EL	29				
SE	49			ES	54		
EE	42	EN	111	AS	41	Total	2,495

TABLE 9–D.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS

CA	20	ND	52	EF	18	SI	34
DA	32	RD	17	OF	25	TI	45
EA	35						
HA	20	BE	18	IG	19	AL	32
LA	28	CE	32	NG	27	EL	29
MA	36	DE	33			IL	23
NA	26	EE	42	CH	14	LL	27
PA	14	GE	14	GH	20	OL	19
RA	39	HE	20	SH	26		
SA	24	IE	13	TH	78		
TA	28	LE	37			AM	14
		ME	26	AI	17	EM	14
AC	14	NE	57	DI	27	OM	25
EC	32	PE	23	EI	27		
IC	22	RE	98	FI	39	AN	64
NC	19	SE	49	HI	33	EN	111
		TE	71	LI	20	IN	75
AD	27	VE	57	NI	30	ON	77
ED	60	WE	22	RI	30	UN	21

TABLE 9–D, Contd.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*——

(1) AND ACCORDING TO THEIR INITIAL LETTERS—Continued

CO	41	AR	44	OS	14	YT	15
DO	16	TR	17	IS	35		
FO	40	UR	31	RS	31	AU	13
HO	20	ER	87			OU	37
IO	41	OR	64	AT	47	QU	15
LO	13	PR	18	CT	14		
NO	18	HR	17	DT	15	EV	20
PO	17	IR	27	ET	37	IV	25
RO	28			HT	28	TW	36
SO	15	AS	41	IT	27		
TO	50	SS	19	NT	82	IX	15
WO	19	TS	19	OT	19		
		DS	13	RT	42	TY	41
EP	20	ES	54	ST	63		
OP	25	NS	24	TT	19	Total	3,745

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES

RA	39	TE	71	TH	78	AM	14
MA	36	NE	57	SH	26	EM	14
EA	35	VE	57	GH	20		
DA	32	SE	49	CH	14	EN	111
LA	28	EE	42			ON	77
TA	28	LE	37	TI	45	IN	75
NA	26	DE	33	FI	39	AN	64
SA	24	CE	32	SI	34	UN	21
CA	20	ME	26	HI	33		
HA	20	PE	23	NI	30	TO	50
PA	14	WE	22	RI	30	CO	41
		HE	20	DI	27	IO	41
EC	32			EI	27	FO	40
IC	22	BE	18	LI	20	RO	28
NC	19	GE	14	AI	17	HO	20
AC	14	IE	13			WO	19
				AL	32	NO	18
ED	60	OF	25	EL	29	PO	17
ND	52	EF	18	LL	27	DO	16
AD	27			IL	23	SO	15
RD	17	NG	27	OL	19	LO	13
RE	98	IG	19	OM	25		

TABLE 9–D, Concluded.—*The 117 digraphs composing 75% of the 5,000 digraphs of Table 6, arranged alphabetically according to their final letters*

(2) AND ACCORDING TO THEIR ABSOLUTE FREQUENCIES—Continued

OP	25	ES	54	AT	47	QU	15
EP	20	AS	41	RT	42	AU	13
		IS	35	ET	37		
		RS	31	HT	28	IV	25
ER	87	NS	24	IT	27	EV	20
OR	64	SS	19	OT	19		
AR	44	TS	19	TT	19	TW	36
UR	31	OS	14	DT	15	IX	15
IR	27	DS	13	YT	15		
PR	18			CT	14	TY	41
HR	17	NT	82				
TR	17	ST	63	OU	37	Total	3,745

TABLE 9–E.—*All the 438 different digraphs of Table 6 arranged alphabetically first according to their final letters and then according to their initial letters*

(SEE TABLE 6.—READ DOWN THE COLUMNS)

TABLE 10–A.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged according to their absolute frequencies*

ENT	569	TOP	174	EIG	135
ION	260	NTH	171	FIV	135
AND	228	TWE	170	MEN	131
ING	226	TWO	163	SEV	131
IVE	225	ATI	160	ERS	126
TIO	221	THR	158	UND	125
FOR	218	NTY	157	NET	118
OUR	211	HRE	153	PER	115
THI	211	WEN	153	STA	115
ONE	210	FOU	152	TER	115
NIN	207	ORT	146	EQU	114
STO	202	REE	146	RED	113
EEN	196	SIX	146	TED	112
GHT	196	ASH	143	ERI	109
INE	192	DAS	140	HIR	106
VEN	190	IGH	140	IRT	105
EVE	177	ERE	138	DER	101
EST	176	COM	136	DRE	100
TEE	174	ATE	135		

TABLE 10–B. *The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their initial letters and then according to their absolute frequencies*

AND	228	GHT	196	REE	146
ATI	160			RED	113
ASH	143	HRE	153		
ATE	135	HIR	106	STO	202
				SIX	146
COM	136	ION	260	SEV	131
		ING	226	STA	115
DAS	140	IVE	225		
DER	101	INE	192	TIO	221
DRE	100	IGH	140	THI	211
		IRT	105	TEE	174
ENT	569			TOP	174
EEN	196	MEN	131	TWE	170
EVE	177			TWO	163
EST	176	NIN	207	THR	158
ERE	138	NTH	171	TER	115
EIG	135	NTY	157	TED	112
ERS	126	NET	118		
EQU	114			UND	125
ERI	109	OUR	211		
		ONE	210	VEN	190
FOR	218	ORT	146		
FOU	152			WEN	153
FIV	135	PER	115		

TABLE 10–C.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their central letters and then according to their absolute frequencies*

DAS	140	DER	101	HIR	106
EEN	196	IGH	140	ENT	569
VEN	190			AND	228
TEE	174	THI	211	ING	226
WEN	153	GHT	196	ONE	210
REE	146	THR	158	INE	192
MEN	131			UND	125
SEV	131	TIO	221		
NET	118			ION	260
PER	115	NIN	207	FOR	218
TER	115	SIX	146	TOP	174
RED	113	EIG	135	FOU	152
TED	112	FIV	135	COM	136

TABLE 10–C, Concluded.—*The 56 trigrams appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their central letters and then according to their absolute frequencies*

EQU	114	DRE	100	STA	115
HRE	153	EST	176	OUR	211
ORT	146	ASH	143	IVE	225
ERE	138	STO	202	EVE	177
ERS	126	NTH	171	TWE	170
ERI	109	ATI	160	TWO	163
IRT	105	NTY	157		
		ATE	135		

TABLE 10–D.—*The 56 trigraphs appearing 100 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their final letters and then according to their absolute frequencies*

STA	115	IGH	140	TER	115
AND	228	THI	211	HIR	108
UND	125	ATI	160	DER	101
RED	113	ERI	109	DAS	140
TED	112	COM	136	ERS	126
IVE	225	ION	260	ENT	569
ONE	210	NIN	207	GHT	196
INE	192	EEN	196	EST	176
EVE	177	VEN	190	ORT	146
TEE	174	WEN	153	NET	118
TWE	170	MEN	131	IRT	105
HRE	153	TIO	221	FOU	152
REE	146	STO	202	EQU	114
ERE	138	TWO	163	FIV	135
ATE	135	TOP	174	SEV	131
DRE	100	FOR	218	SIX	146
ING	226	OUR	211	NTY	157
EIG	135	THR	158		
NTH	171	PER	115		
ASH	143				

TABLE 11-A.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged according to their absolute frequencies*

TION	218	THIR	104	ASHT	64
EVEN	168	EENT	102	HUND	64
TEEN	163	REQU	98	DRED	63
ENTY	161	HIRT	97	RIOD	63
STOP	154	COMM	93	IVED	62
WENT	153	QUES	87	ENTS	62
NINE	153	UEST	87	FFIC	62
TWEN	152	EQUE	86	FROM	59
THRE	149	NDRE	77	IRTY	59
FOUR	144	OMMA	71	RTEE	59
IGHT	140	LLAR	71	UNDR	59
FIVE	135	OLLA	70	NAUG	56
HREE	134	VENT	70	OURT	56
EIGH	132	DOLL	68	UGHT	56
DASH	132	LARS	68	STAT	54
SEVE	121	THIS	68	AUGH	52
ENTH	114	PERI	67	CENT	52
MENT	111	ERIO	66	FICE	50

TABLE 11-B.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams, arranged first alphabetically according to their initial letters, and then according to their absolute frequencies*

ASHT	64	HREE	134	REQU	98
AUGH	52	HIRT	97	RIOD	63
		HUND	64	RTEE	59
COMM	93				
CENT	52	IGHT	140	STOP	154
		IVED	62	SEVE	121
DASH	132	IRTY	59	STAT	54
DOLL	68				
DRED	63	LLAR	71	TION	218
		LARS	68	TEEN	163
EVEN	168			TWEN	152
ENTY	161	MENT	111	THRE	149
EIGH	132			THIR	104
ENTH	114	NINE	153	THIS	68
EENT	102	NDRE	77		
EQUE	86	NAUG	56	UEST	87
ERIO	66			UNDR	59
ENTS	62	OMMA	71	UGHT	56
		OLLA	70		
FOUR	144	OURT	56	VENT	70
FIVE	135				
FFIC	62	PERI	67	WENT	153
FROM	59				
FICE	50	QUES	87		

TABLE 11–C.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their second letters and then according to their absolute frequencies*

DASH	132	THIS	68	EQUE	86
LARS	68				
NAUG	56	TION	218	HREE	134
		NINE	153	ERIO	66
NDRE	77	FIVE	135	DRED	63
		EIGH	132	FROM	59
TEEN	163	HIRT	97	IRTY	59
WENT	153	RIOD	63		
SEVE	121	FICE	50	ASHT	64
MENT	111				
EENT	102	LLAR	71	STOP	154
REQU	98	OLLA	70	RTEE	59
UEST	87			STAT	54
VENT	70	OMMA	71		
PERI	67			QUES	87
CENT	52	ENTY	161	HUND	64
		ENTH	114	CURT	56
FFIC	62	ENTS	62	AUGH	52
		UNDR	59		
IGHT	140			EVEN	168
UGHT	56	FOUR	144	IVED	62
		COMM	93		
THRE	149	DOLL	68	TWEN	152
THIR	104				

TABLE 11–D.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their third letters and then according to their absolute frequencies*

LLAR	71	EIGH	132	COMM	93
STAT	54	AUGH	52	OMMA	71
FICE	50	IGHT	140	WENT	153
		ASHT	64	NINE	153
UNDR	59	UGHT	56	MENT	111
				EENT	102
EVEN	168	THIR	104	VENT	70
TEEN	163	THIS	68	HUND	64
TWEN	152	ERIO	66	CENT	52
HREE	134	FFIC	62		
QUES	87			TION	218
DRED	63	OLLA	70	STOP	154
IVED	62	DOLL	68	RIOD	63
RTEE	59			FROM	59

TABLE 11–D, Concluded.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their third letters and then according to their absolute frequencies*

REQU	98	OURT	56	IRTY	59
		DASH	132	FOUR	144
THRE	149	UEST	87	EQUE	86
HIRT	97			NAUG	56
NDRE	77	ENTY	161		
LARS	68	ENTH	114	FIVE	135
PERI	67	ENTS	62	SEVE	121

TABLE 11–E.—*The 54 tetragraphs appearing 50 or more times in the 50,000 letters of Government plain-text telegrams arranged first alphabetically according to their final letters and then according to their absolute frequencies*

OMMA	71	DASH	132	QUES	87
OLLA	70	EIGH	132	THIS	68
		ENTH	114	LARS	68
		AUGH	52	ENTS	62
FFIC	62				
		PERI	67		
				WENT	153
HUND	64	DOLL	68	IGHT	140
DRED	63			MENT	111
RIOD	63	COMM	93	EENT	102
IVED	62	FROM	59	HIRT	97
				UEST	87
		TION	218	VENT	70
NINE	153	EVEN	168	ASHT	64
THRE	149	TEEN	163	UGHT	56
FIVE	135	TWEN	152	OURT	56
HREE	134			STAT	54
SEVE	121	ERIO	66	CENT	52
EQUE	86				
NDRE	77	STOP	154		
RTEE	59			REQU	98
FICE	50	FOUR	144		
		THIR	104		
		LLAR	71	ENTY	161
NAUG	56	UNDR	59	IRTY	59

TABLE 12.—*Average and mean lengths of words*

Number of letters in word	Number of times word appears	Number of letters
1	378	378
2	973	1, 946
3	1, 307	3, 921
4	1, 635	6, 540
5	1, 410	7, 050
6	1, 143	6, 858
7	1, 009	7, 063
8	717	5, 736
9	476	4, 284
10	274	2, 740
11	161	1, 771
12	86	1, 032
13	23	299
14	23	322
15	4	60
120	9, 619	50, 000

(1) Mean length of messages ..... 5.2 Letters.
(2) Average length of messages ..... 217 Letters.
(3) Mean length of messages ..... 191 Letters.
(4) Mode (most frequent) length ..... 105–114 Letters.
(5) It is extremely unusual to find 5 consecutive letters without at least one vowel.
(6) The average number of letters between vowels is 2.

INDEX

INDEX

	Paragraphs	Pages
Accented letters	5b	8.
Alphabets:		
Bipartite	35c	59.
Deciphering	31c	52.
Direct standard	12a, 16, 19	18, 26, 31–33.
Enciphering	29b, 31c	49, 52.
Keyword-mixed	31d	53.
Mixed	12a, 15a, 19, 21d, 22b, 24c, 31b.	18, 25, 31–33, 39, 39, 41, 52.
Reversed standard	12a, 16, 19b, 20b	18, 26, 33, 36.
Standard	12a, 15a, 16, 19, 20b, 23, 38e.	18, 25, 26, 31–33, 36, 40, 65.
Systematically mixed	31c, e	52, 53.
Analytic key for cryptanalysis	6d, 50	9, 103–104.
Arbitrary symbols	13h, 48	22, 100–101.
Assumptions	46h	98.
Average length of messages	11b	16.
Baconian cipher	35e	60.
Beginnings of messages	32e	54.
Biliteral substitution	41	70–71.
Bipartite alphabet	35b, c	59.
Blanks, number of	14e	24.
Book systems	49e	102
Censorship, methods for evading	47c	99.
Characteristic frequency of the letters of a language	9d, 14b, 25	12, 23, 41.
Characteristic frequency, suppression of	37, 41f	63, 71.
Checkerboard systems	44, 45	73–83, 83.
Checkerboards, 4-square	44	73–83.
Chinese Official Telegraph Code	48e	101.
Cipher:		
Baconian	35e	60,
Component	34	57–58.
Distinguished from code	6c, 38c	9, 64.
Text, length of, as compared with plain text	40c	69.
Unit	41c	70.
Classification of ciphers	12a, 13, 47, 50e, f	18, 18–22, 99, 104, 104.
Code systems	4a, 41g, 47b, 48d, e	7, 71, 99, 100, 101.
Distinguished from cipher	6c, 38c	9, 64.
Completing the plain component	20a, 34a	34, 57.
Concealed messages	47c	99.
Condensed table of repetitions	27i	46.
Consonants:		
Distinguished from vowels	28, 32c	46–47, 53.
Relative frequency of	10a, 13, 19	13, 18–22, 31–33.
In succession	32c	53.
Conversion of cipher text	21a, c, 34c	38, 38, 58.
Coordinates on work sheet	26d	42.

	Paragraphs	Pages
Coordination of services	2*e*	4.
Crests and troughs	10*a*, 14*b*, 41*f*, 44*c*	13, 23, 71, 74.
Absence of	14*c*	24.
Deciphering alphabets	31*c*	52.
Dictionary words used as code words	47*b*	99.
Digraphic substitution	41*c*, 42, 43	70, 72, 72–73.
Digraphs:		
Characteristic frequency	25	41.
Weighted according to relative frequency	29	48–49.
Distribution:		
Frequency	9*a*, 11*b*	11, 16.
Normal	17*b*, *c*	27, 28.
With no crests and troughs	14*c*	24
Dummy letters	47*c*	99.
Elementary sounds, characteristic frequency of	14*b*	23.
Enciphering alphabet	31*c*	52.
Endings of messages	32*e*	54.
Equivalent values	39*b*	66.
Figure ciphers	13*h*, 48	22, 100–101.
Fitting distribution to normal	17*b*, *c*, 19, 38*e*	27, 28, 31–33, 65.
Foreign language cryptograms	5*b*, *c*	8.
Formulas	33*d*	56.
Frequency distribution	9, 17, 19, 26*e*, 44*c*	11–13, 27–28, 31–33, 42, 74.
Fitted to normal	17*b*	27, 28.
For certain types of code	47*b*	99.
Four part	38*d*	65.
Multiliteral	35, 37, 41*c*	59–60, 63, 70.
Uniliteral	9, 17	11–13, 27–28.
Triliteral	27	43–46.
Frequency method of solution	18, 24*d*, 29	29–31, 41, 48–49.
General solutions in cryptanalysis	46*i*	98.
General system, determination of	4*a*, 6, 13, 50	7, 8–9, 18–22, 103–104.
Generatrix	20*a*	34.
Goodness of fit	17*b*	27.
Grilles	47*c*	99.
Hidden messages	47*c*	99.
High-frequency consonants	18*d*	19.
Historical examples of multiliteral systems	36	61.
Idiomorphism	33*e*	56.
Indicators	49*b*	101.
Intelligence facilities	2*e*	4.
Intelligible text obtained by chance	21*b*	38.
Intuitive method	33	55–57.
Invisible writing	1*d*	1.
Japanese Morse alphabet	5*b*, 48*e*	8, 101.
Kata Kana Morse alphabet	48*e*	101.
Key, analytical	6*d*, 50	9, 103–104.
Known sequences	23	40.
Language employed in a cryptogram	4*a*, 5	7, 8.
Language frequency characteristics	9*d*, 25	12, 41.
Language peculiarities	5*b*	8.

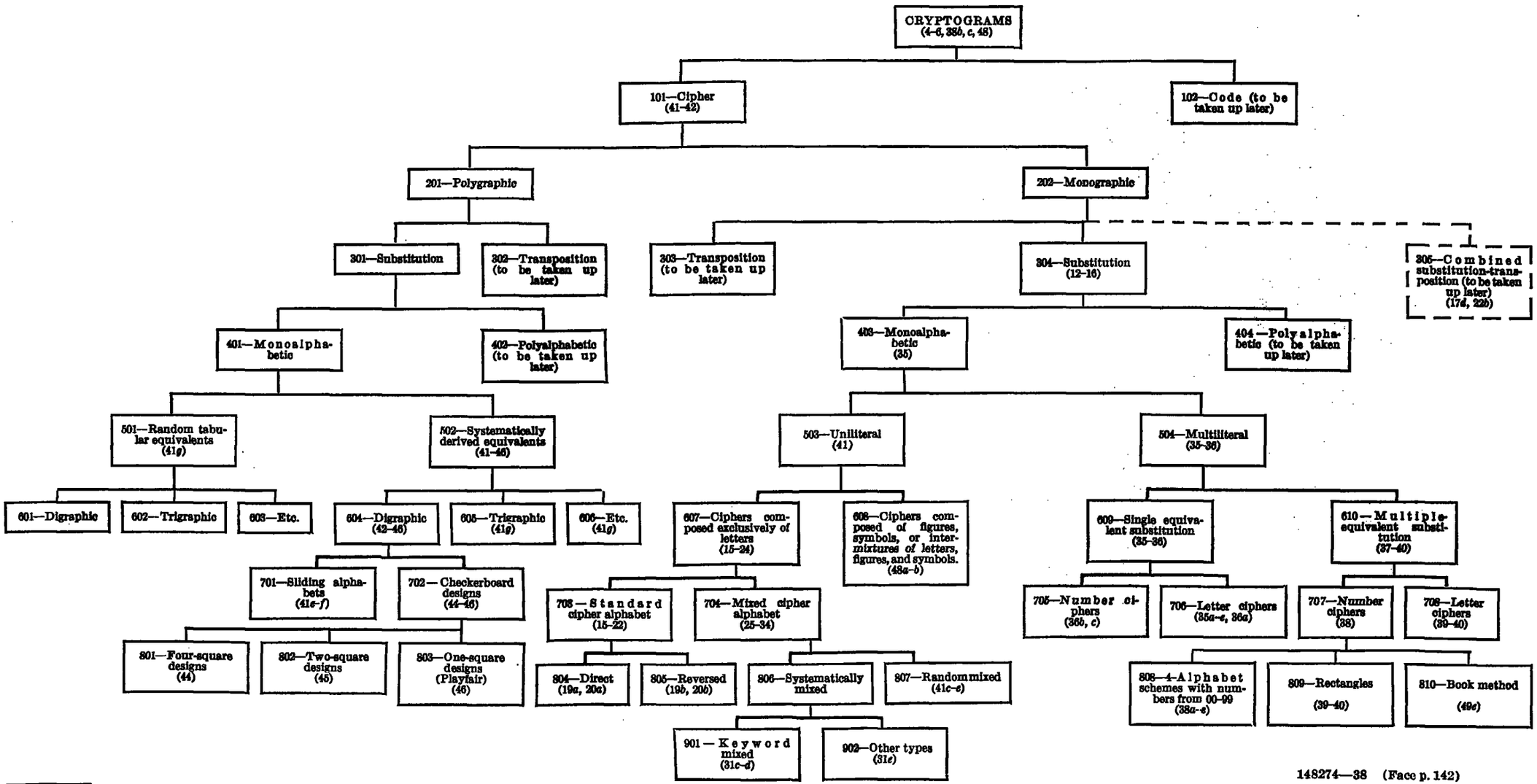
	Paragraphs	Pages
Letters:		
Accented	5b	8.
Low-frequency	31c	52.
Missing	5b, 14e	8, 24.
Low-frequency consonants	13d, 31c	19, 52.
Medium-frequency consonants	13d	19.
Messages:		
Beginnings and endings amenable to cryptanalysis	32e	54.
General phraseology	49a	101.
Hidden	47c	99.
Military text	10b	15.
Missing letters	5b, 14e	8, 24.
Mixed alphabet	15a, 22b, 24c, 31b	25, 39, 41, 52.
Mixed sequence	21d	39.
Modified Playfair	46d	86.
Monoalphabets	1b	1.
Monoalphabet distinguished from polyalphabet	12, 14	18, 22–25.
Morse alphabet, Japanese, Russian	5b, 48e	8, 101.
Multiliteral substitution	35, 37, 41c	59–60, 63, 70.
Multiliteral systems, historical examples of	36	61.
Normal distribution	17b, c	27, 28.
Normal frequency	9, 11, 25	11–13, 16–17, 41.
Deviations from	13b	19.
Nulls	40, 47c	68–69, 99.
New York Tribune, ciphers in	36	61.
Patterns	33d	56.
Phraseology of messages	49a	101.
Plain component, completion of	20a	34.
Plain-text unit	41c	70.
Playfair cipher	44, 46	73–83, 84–98.
Modified	46d	86.
Polyalphabetic cipher distinguished from monoalphabet	12, 14	18, 22–25.
Polygraphic substitution	41	70–71.
Prefixes in triliteral distribution	27e	44.
Prerequisites for cryptographic work	2	2–5.
Probable-word method	33	55–57.
Pseudo-polygraphic systems	41e	71.
Punctuation in telegraphic text	10c	15.
Quadraliteral cipher	35d	60.
Quinqueliteral cipher	35e	60.
Random text, number of blanks	14f	24.
Relative frequencies	10b, c, d, 11, 14b	15, 15, 15, 16–17, 23.
Repetitions	13g, 24b, c, 27	21, 40, 41, 43–46.
In a code message	38c	64.
Of consonants	32c	53.
Of digraphs and trigraphs	27f	46.
Condensed table of	27i	46.
Reversed standard alphabets	16, 20b	26, 36.
Reversible digraphs indicated on worksheet	26f	43.
Russian Morse alphabet	5b, 48e	8, 101.
Security of monoalphabet using standard alphabets	23	40.
Sequences:		
Known	23	40.
Mixed	21d	39.
Unknown	23	40.

148274—38——10

	Paragraphs	Pages
Solutions of a subjective nature	3	5.
Specific key	4, 7, 19*a*, 31*b*	7, 10, 31, 52.
Standard alphabets	15*a*, 16, 20*b*, 38*e*	25, 26, 36, 65.
Subjective solutions	3	5.
Substitution:		
Biliteral	41	70–71.
Digraphic	41*c*, 42*a*	70, 72.
Distinguished from transposition	12, 13	18, 18–22.
Polygraphic	41*c*	70.
Multiliteral	41*c*	70.
Trigraphic	41*c*	70.
Triliteral	41	70–71.
Suffixes in trigraphic distribution	27*e*	44.
Suppression of frequency	37, 40*b*, 41*f*	63, 69, 71.
Symbols as cipher elements	13*h*, 48	22, 100–101.
Telegrams, average length of	11*b*	16.
Terminology	1	1.
Text, different types of	10*b*, *c*	15, 15.
Transposition distinguished from substitution	12, 13	18, 18–22.
Trigraphic cipher system	41*c*	70.
Trigraphic frequency table	27 (footnote 2)	44.
Triliteral frequency distribution	27	43–46.
Triliteral substitution	41	70–71.
Type numbers for cryptographic systems	50*f*	104.
Uniliteral frequency distribution	9, 17	11–13, 27–28.
Unknown sequences	23	40.
Variants	37*d*, 40*b*, 49*c*, *d*, *f*	63, 69, 101, 102, 102.
Vowels:		
Average distance apart	32*c*, footnote	53.
Combinations with consonants	28, 29	46–47, 48–49.
Combinations with vowels	29*a*	48.
Distinguished from consonants	28, 32*c*	46–47, 53.
In succession	32*c*	53.
Relative frequency of	10*a*, 13, 19	13, 18–22, 31–33.
Word formulas	33*d*	56.
Word lengths in a cryptogram	26*c*, 32*e*, 33*d*, *f*, *g*	42, 54, 56, 56, 57.
Word patterns	33*d*	56.
Word skeletons	30*b*, 32*e*	49, 54.
Work sheet, preparation of	26	42–43.

ANALYTICAL KEY FOR CRYPTANALYSIS [1]

(Numbers in parentheses refer to paragraph numbers in this text)



148274—38 (Face p. 142)

[1] For explanation, see Par. 50.

[redacted]

~~Confidential~~

WAR DEPARTMENT

OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

MILITARY CRYPTANALYSIS

PART II

[redacted]

William F. Friedman
Washington
1938

30 April 1959

~~This document is re-graded "CONFIDENTIAL" UP of DOD Directive 5200.1 dated 8 July 1957, and by authority of the Director, National Security Agency.~~

Paul S. Willard
Paul S. Willard
Colonel, AGC
Adjutant General

~~Restricted~~

WAR DEPARTMENT

OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

MILITARY CRYPTANALYSIS
Part II

SIMPLER VARIETIES
OF POLYALPHABETIC SUBSTITUTION SYSTEMS

By

WILLIAM F. FRIEDMAN

Principal Cryptanalyst
Chief of Signal Intelligence Section
War Plans and Training Division

PREPARED UNDER THE DIRECTION OF THE
CHIEF SIGNAL OFFICER

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1938

MILITARY CRYPTANALYSIS. PART II. SIMPLER VARIETIES OF POLYALPHABETIC SUBSTITUTION SYSTEMS

Section		Paragraphs	Pages
I.	Introductory remarks	1–4	1–3
II.	Cipher alphabets for polyalphabetic substitution	5–7	4–9
III.	Theory of solution of repeating-key systems	8–12	10–16
IV.	Repeating-key systems with standard cipher alphabets	13–15	17–23
V.	Repeating-key systems with mixed cipher alphabets, I	16–26	24–48
VI.	Repeating-key systems with mixed cipher alphabets, II	27–30	49–51
VII.	Theory of indirect symmetry of position in secondary alphabets	31	52–59
VIII.	Application of principles of indirect symmetry of position	32–36	60–77
IX.	Repeating-key systems with mixed cipher alphabets, III	37–40	78–83
X.	Repeating-key systems with mixed cipher alphabets, IV	41–46	84–95
	Appendix 1		96–107
	Appendix 2		108–118
	Index		119–120

SECTION I

INTRODUCTORY REMARKS

	Paragraph
The essential difference between monoalphabetic and polyalphabetic substitution	1
Primary classification of polyalphabetic systems	2
Primary classification of periodic systems	3
Sequence of study of polyalphabetic systems	4

1. The essential difference between monoalphabetic and polyalphabetic substitution.—*a.* In the substitution methods thus far discussed it has been pointed out that their basic feature is that of monoalphabeticity. From the cryptanalytic standpoint, neither the nature of the cipher symbols, nor their method of production is an essential feature, although these may be differentiating characteristics from the cryptographic standpoint. It is true that in those cases designated as monoalphabetic substitution with variants or multiple equivalents, there is a departure, more or less considerable, from strict monoalphabeticity. In some of those cases, indeed, there may be available two or more wholly independent sets of equivalents, which, moreover, may even be arranged in the form of completely separate alphabets. Thus, while a loose terminology might permit one to designate such systems as polyalphabetic, it is better to reserve this nomenclature for those cases wherein polyalphabeticity is the essence of the method, specifically introduced with the purpose of imparting a *positional* variation in the substitutive equivalents for plain-text letters, in accordance with some rule directly or indirectly connected with the absolute *positions* the plain-text letters occupy in the message. This point calls for amplification.

b. In monoalphabetic substitution with variants the object of having different or multiple equivalents is to suppress, so far as possible by simple methods, the characteristic frequencies of the letters occurring in plain text. As has been noted, it is by means of these characteristic frequencies that the cipher equivalents can usually be identified. In these systems the varying equivalents for plain-text letters are subject to the free choice and caprice of the enciphering clerk; if he is careful and conscientious in the work, he will really make use of all the different equivalents afforded by the system; but if he is slip-shod and hurried in his work, he will use the same equivalents repeatedly rather than take pains and time to refer to the charts, tables, or diagrams to find the variants. Moreover, and this is a crucial point, even if the individual enciphering clerks are extremely careful, when many of them employ the same system it is entirely impossible to insure a complete diversity in the encipherments produced by two or more clerks working at different message centers. The result is inevitably to produce plenty of repetitions in the texts emanating from several stations, and when texts such as these are all available for study they are open to solution, by a comparison of their similarities and differences.

c. In true polyalphabetic systems, on the other hand, there is established a rather definite procedure which automatically determines the shifts or changes in equivalents or in the manner in which they are introduced, so that these changes are beyond the momentary whim or choice of the enciphering clerk. When the method of shifting or changing the equivalents is scientifically sound and sufficiently complex, the research necessary to establish the values of the cipher characters is much more prolonged and difficult than is the case even in complicated monoalphabetic substitution with variants, as will later be seen. These are the objects of true polyalphabetic substitution systems. The number of such systems is quite large, and it will be possible to

describe in detail the cryptanalysis of only a few of the more common or typical examples of methods encountered in practical military communications.

d. The three methods, (1) single-equivalent monoalphabetic substitution, (2) monoalphabetic substitution with variants, and (3) true polyalphabetic substitution, show the following relationships as regards the equivalency between plain-text and cipher-text units:

A. In method (1), there is a set of 26 symbols; a plain-text letter is always represented by one and only one of these symbols; conversely, a symbol always represents the same plain-text letter. The equivalence between the plain-text and the cipher letters is constant in both encipherment and decipherment.

B. In method (2), there is a set of n symbols, where n may be any number greater than 26 and often is a multiple of that number; a plain-text letter may be represented by 1, 2, 3, . . . different symbols; conversely, a symbol always represents the same plain-text letter, the same as is the case in method (1). The equivalence between the plain-text and the cipher letters is variable in encipherment but constant in decipherment.[1]

C. In method (3) there is, as in the first method, a set of 26 symbols; a plain-text letter may be represented by 1, 2, 3, . . . 26 different symbols; conversely, a symbol may represent 1, 2, 3, . . . 26 different plain text letters, depending upon the system and the specific key. The equivalence between the plain-text and the cipher letters is variable in both encipherment and decipherment.

2. Primary classification of polyalphabetic systems.—*a.* A primary classification of polyalphabetic systems into two rather distinct types may be made: (1) periodic systems and (2) aperiodic systems. When the enciphering process involves a cryptographic treatment which is repetitive in character, and which results in the production of *cyclic phenomena* in the cryptographic text, the system is termed *periodic.* When the enciphering process is not of the type described in the foregoing general terms, the system is termed *aperiodic.* The substitution in both cases involves the use of two or more cipher alphabets.

b. The cyclic phenomena inherent in a periodic system may be exhibited externally, in which case they are said to be *patent*, or they may not be exhibited externally, and must be uncovered by a preliminary step in the analysis, in which case they are said to be *latent.* The periodicity may be quite definite in nature, and therefore determinable with mathematical exactitude allowing for no variability, in which case the periodicity is said to be *fixed.* In other instances the periodicity is more or less flexible in character and even though it may be deter-

[1] There is a monoalphabetic method in which the inverse result obtains, the correspondence being constant in encipherment but variable in decipherment; this is a method not found in the usual books on cryptography but in an essay on that subject by Edgar Allan Poe, entitled, in some editions of his works, *A few words on secret writing* and, in other editions, *Cryptography.* The method is to draw up an enciphering alphabet such as the following (using Poe's example):

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	S	U	A	V	I	T	E	R	I	N	M	O	D	O	F	O	R	T	I	T	E	R	I	N	R	E

In such an alphabet, because of repetitions in the cipher component, the plain-text equivalents are subject to a considerable degree of variability, as will be seen in the deciphering alphabet:

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	C			M	G	O		E			K	J	L				H	A	F	B	D					
					U			I				X	N				Q		R							
					Z			S					P				V		T							
								W									Y									

This type of variability gives rise to ambiguities in decipherment. A cipher group such as TIE, would yield such plain-text sequences as REG, FIG, TEU, REU, etc., which could be read only by *context.* No system of such a character would be practical for serious usage. For a further discussion of this type of cipher alphabet see Friedman, William F., *Edgar Allan Poe, Cryptographer,* Signal Corps Bulletins Nos. 97 and 98, 1937–38.

minable mathematically, allowance must be made for a degree of variability subject to limits controlled by the specific system under investigation. The periodicity is in this case said to be *flexible*, or *variable within limits*.

3. Primary classification of periodic systems.—*a*. Periodic polyalphabetic substitution systems may primarily be classified into two kinds:

(1) Those in which only a few of a whole set of cipher alphabets are used in enciphering individual messages, these alphabets being employed repeatedly in a fixed sequence throughout each message. Because it is usual to employ a secret word, phrase, or number as a key to determine the number, identity, and sequence with which the cipher alphabets are employed, and this key is used over and over again in encipherment, this method is often called the *repeating-key system*, or the *repeating-alphabet system*. It is also sometimes referred to as the *multiple-alphabet system* because if the keying of the entire message be considered as a whole it is composed of multiples of a short key used repetitively.[2] In this text the designation "repeating-key system" will be used.

(2) Those in which all the cipher alphabets comprising the complete set for the system are employed one after the other successively in the encipherment of a message, and when the last alphabet of the series has been used, the encipherer begins over again with the first alphabet. This is commonly referred to as a *progressive-alphabet system* because the cipher alphabets are used in progression.

4. Sequence of study of polyalphabetic systems.—*a*. In the studies to be followed in connection with polyalphabetic systems, the order in which the work will proceed conforms very closely to the classifications made in paragraphs 2 and 3. Periodic polyalphabetic substitution ciphers will come first, because they are, as a rule, the simpler and because a thorough understanding of the principles of their analysis is prerequisite to a comprehension of how aperiodic systems are solved. But in the final analysis the solution of examples of both types rests upon the conversion or reduction of polyalphabeticity into monoalphabeticity. If this is possible, solution can always be achieved, granted there are sufficient data in the final monoalphabetic distributions to permit of solution by recourse to the ordinary principles of frequency.

b. First in the order of study of periodic systems will come the analysis of repeating-key systems. Some of the more simple varieties will be discussed in detail, with examples. Subsequently, ciphers of the progressive type will be discussed. There will then follow a more or less detailed treatment of aperiodic systems.

---

[2] French terminology calls this the "double-key method", but there is no logic in such nomenclature.

SECTION II

CIPHER ALPHABETS FOR POLYALPHABETIC SUBSTITUTION

	Paragraph
Classification of cipher alphabets upon the basis of their derivation	5
Primary components and secondary alphabets	6
Primary components, cipher disks, and square tables	7

5. Classification of cipher alphabets upon the basis of their derivation.—*a.* The substitution processes in polyalphabetic methods involve the use of a plurality of cipher alphabets. The latter may be derived by various schemes, the exact nature of which determines the principal characteristics of the cipher alphabets and plays a very important role in the preparation and solution of polyalphabetic cryptograms. For these reasons it is advisable, before proceeding to a discussion of the principles and methods of analysis, to point out these various types of cipher alphabets, show how they are produced, and how the method of their production or derivation may be made to yield important clues and short-cuts in analysis.

b. A primary classification of cipher alphabets for polyalphabetic substitution may be made into the two following types:

(1) Independent or unrelated cipher alphabets.

(2) Derived or interrelated cipher alphabets.

c. Independent cipher alphabets may be disposed of in a very few words. They are merely separate and distinct alphabets showing no relationship to one another in any way. They may be compiled by the various methods discussed in Section IX of *Elementary Military Cryptography*. The solution of cryptograms written by means of such alphabets is rendered more difficult by reason of the absence of any relationship between the equivalents of one cipher alphabet and those of any of the other alphabets of the same cryptogram. On the other hand, from the point of view of practicability in their production and their handling in cryptographing and decryptographing, they present some difficulties which make them less favored by cryptographers than cipher alphabets of the second type.

d. Derived or interrelated alphabets, as their name indicates, are most commonly produced by the *interaction of* two primary components, which when juxtaposed at the various points of coincidence can be made to yield *secondary alphabets.*[1]

6. Primary components and secondary alphabets.—Two basic, slidable sequences or components of n characters each will yield n secondary alphabets. The components may be classified according to various schemes. For cryptanalytic purposes the following classification will be found useful:

CASE A. The primary components are both normal sequences.

(1) The sequences proceed in the same direction. (The secondary alphabets are direct standard alphabets.) (Pars. 13–15.)

(2) The sequences proceed in opposite directions. (The secondary alphabets are reversed standard alphabets; they are also reciprocal cipher alphabets.) (Par. 13*i*, 14*g*.)

CASE B. The primary components are not both normal sequences.

(1) The plain component is normal, the cipher component is a mixed sequence. (The secondary alphabets are mixed alphabets.) (Par. 16–25.)

[1] See Sec. VIII and IX, *Elementary Military Cryptography.*

(2) The plain component is a mixed sequence, the cipher component is normal. (The secondary alphabets are mixed alphabets.) (Par. 26.)

(3) Both components are mixed sequences.

(a) Components are identical mixed sequences.

I. Sequences proceed in the same direction. (The secondary alphabets are mixed alphabets.) (Par. 28.)

II. Sequences proceed in opposite directions. (The secondary alphabets are reciprocal mixed alphabets.) (Par. 38.)

(b) Components are different mixed sequences. (The secondary alphabets are mixed alphabets.) (Par. 39.)

7. Primary components, cipher disks, and square tables.—*a.* In preceding texts it has been shown that the equivalents obtainable from the use of quadricular or square tables may be duplicated by the use of revolving cipher disks or of sliding primary components. It was also stated that there are various ways of employing such tables, disks, and sliding components. Cryptographically the results may be quite diverse from different methods of using such paraphernalia, since the specific equivalents obtained from one method may be altogether different from those obtained from another method. But from the cryptanalytic point of view the diversity referred to is of little significance; only in one or two cases does the specific method of employing these cryptographic instrumentalities have an important bearing upon the procedure in cryptanalysis. However, it is advisable that the student learn something about these different methods before proceeding with further work.

b. There are, not *two*, but *four* letters involved in every case of finding equivalents by means of sliding primary components; furthermore, the determination of an equivalent for a given plain-text letter is representable by *two* equations involving *four* elements, usually letters. Three of these letters are by this time well-known to and understood by the student, viz, θ_k, θ_p, and θ_c. The fourth element or letter has been passed over without much comment, but cryptographically it is just as important a factor as the other three. Its function may best be indicated by noting what happens when two primary components are juxtaposed, for the purpose of finding equivalents. Suppose these components are the following sequences:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Now suppose one is merely asked to find the equivalent of P_p when the key letter is K. Without further specification, the cipher equivalent cannot be stated; for it is necessary to know not only which K will be used as the key letter, the one in the component labeled (1) or the one in the component labeled (2), but also what letter the K_k will be set against, in order to juxtapose the two components. Most of the time, in preceding texts, these two factors have been tacitly assumed to be fixed and well understood: the K_k is sought in the mixed, or cipher component, and this K is set against A in the normal, or plain component. Thus:

```
                          Plain        Index
                            ↓            ↓
(1) Plain.....ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
(2) Cipher.....      FBPYRCQZIGSEHTDJUMKVALWNOX
                          ↑            ↑
                       Cipher        Key
```

With this setting $P_p = Z_c$.

c. The letter A in this case may be termed the *index letter*, symbolized A_i. The index letter constitutes the fourth element involved in the two equations applicable to the finding of equivalents by sliding components. The four elements are therefore these:

(1) The key letter, θ_k
(2) The index letter, θ_i
(3) The plain-text letter, θ_p
(4) The cipher letter, θ_c

The index letter is commonly the initial letter of the component; but this, too, is only a convention. It *might be any letter* of the sequence constituting the component, as agreed upon by the correspondents. *However, in the subsequent discussion it will be assumed that the index letter is the initial letter of the component in which it is located, unless otherwise stated.*

d. In the foregoing case the enciphering equations are as follows:

(I) $K_k=A_i$; $P_p=Z_c$

But there is nothing about the use of sliding components which excludes other methods of finding equivalents than that shown above. For instance, despite the labeling of the two components as shown above, there is nothing to prevent one from seeking the plain-text letter in the component labeled (2), that is, the cipher component, and taking as its cipher equivalent the letter opposite it in the other component labeled (1). Thus:

Cipher ↓ Index ↓
(1) ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
(2) FBPYRCQZIGSEHTDJUMKVALWNOX
↑ Plain ↑ Key

Thus:

(II) $K_k=A_i$; $P_p=K_c$

e. Since equations (I) and (II) yield different resultants, even with the same index, key, and plain-text letters, it is obvious that an accurate formula to cover a specific pair of enciphering equations must include data showing in what component each of the four letters comprising the equations is located. Thus, equations (I) and (II) should read:

(I) K_k in component (2)=A_i in component (1); P_p in component (1)=Z_c in component (2).
(II) K_k in component (2)=A_i in component (1); P_p in component (2)= K_c in component (1).

For the sake of brevity, the following notation will be used:

(1) $K_{k/2}=A_{i/1}$; $P_{p/1}=Z_{c/2}$
(2) $K_{k/2}=A_{i/1}$; $P_{p/2}=K_{c/1}$

f. Employing two sliding components and the four letters entering into an enciphering equation, there are, in all, twelve different resultants possible for the same set of components and the same set of four basic elements. These twelve differences in resultants arise from a set of twelve different enciphering conditions, as set forth below (the notation adopted in subparagraph *e* is used):

(1) $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/1}=\theta_{c/2}$
(2) $\theta_{k/2}=\theta_{i/1}$; $\theta_{p/2}=\theta_{c/1}$
(3) $\theta_{k/1}=\theta_{i/2}$; $\theta_{p/1}=\theta_{c/2}$
(4) $\theta_{k/1}=\theta_{i/2}$; $\theta_{p/2}=\theta_{c/1}$
(5) $\theta_{k/2}=\theta_{p/1}$; $\theta_{i/1}=\theta_{c/2}$
(6) $\theta_{k/2}=\theta_{c/1}$; $\theta_{i/1}=\theta_{p/2}$
(7) $\theta_{k/2}=\theta_{p/1}$; $\theta_{i/2}=\theta_{c/1}$
(8) $\theta_{k/2}=\theta_{c/1}$; $\theta_{i/2}=\theta_{p/1}$
(9) $\theta_{k/1}=\theta_{p/2}$; $\theta_{i/1}=\theta_{c/2}$
(10) $\theta_{k/1}=\theta_{c/2}$; $\theta_{i/1}=\theta_{p/2}$
(11) $\theta_{k/1}=\theta_{p/2}$; $\theta_{i/2}=\theta_{c/1}$
(12) $\theta_{k/1}=\theta_{c/2}$; $\theta_{i/2}=\theta_{p/1}$

g. The twelve resultants obtainable from juxtaposing sliding components as indicated under the preceding subparagraph may also be obtained either from one square table, in which case twelve different methods of finding equivalents must be applied, or from twelve different square tables, in which case one standard method of finding equivalents will serve all purposes.

h. If but one table such as that shown below as Table I–A is employed, the various methods of finding equivalents are difficult to keep in mind.

TABLE I–A

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X
B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F
P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B
Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P
R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y
C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R
Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C
Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q
I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z
G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I
S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G
E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S
H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E
T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H
D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T
J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D
U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J
M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U
K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M
V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K
A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V
L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A
W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L
N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W
O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N
X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O

For example:

(1) For enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/1}=\Theta_{e/2}$:

Locate Θ_p in top sequence; locate Θ_k in first column;

Θ_e is letter within the square at intersection of the two lines thus determined.

Thus:

$$K_{k/2}=A_{1/1};\ P_{p/1}=Z_{e/2}$$

(2) For enciphering equations $\Theta_{k/2}=\Theta_{1/1}$; $\Theta_{p/2}=\Theta_{c/1}$:
Locate Θ_k in first column; follow line to right to Θ_p; proceed up this column; Θ_c is letter at top.
Thus:

$$K_{k/2}=A_{1/1};\ P_{p/2}=K_{c/1}$$

(3) For enciphering equations $\Theta_{k/1}=\Theta_{1/2}$; $\Theta_{p/1}=\Theta_{c/2}$:
Locate Θ_k in top sequence and proceed down column to Θ_1;
Locate Θ_p in top sequence; Θ_c is letter at other corner of rectangle thus formed.
Thus:

$$K_{k/1}=A_{1/2};\ P_{p/1}=X_{c/2}$$

Only three different methods have been shown and the student no doubt already has encountered difficulty in keeping them segregated in his mind. It would obviously be very confusing to try to remember all twelve methods. But if one standard or fixed method of finding equivalents is followed with several different tables, then this difficulty disappears. Suppose that the following method is adopted: Arrange the square so that the plain-text letter may be sought in a separate sequence, arranged alphabetically, above the square and so that the key letter may be sought in a separate sequence, also arranged alphabetically, to the left of the square; look for the plain-text letter in the top row; locate the key letter in the 1st column to the left; find the letter standing within the square at the intersection of the vertical and horizontal lines thus determined. Then *twelve* squares, equivalent to the twelve different conditions listed in subparagraph *f*, can readily be constructed. They are all shown in Appendix 1, pp. 96–107.

i. When these square tables are examined carefully, certain interesting points are noted. In the first place, the tables may be paired so that one of a pair may serve for enciphering and the other of the pair may serve for deciphering, or vice versa. For example, tables I and II bear this reciprocal relationship to each other; III and IV, V and VI, VII and VIII, IX and X, XI and XII. In the second place, the internal dispositions of the letters, although the tables are derived from the same pair of components, are quite diverse. For example, in table I–B the horizontal sequences are identical, but are merely displaced to the right and to the left different intervals according to the successive key letters. Hence this square shows what may be termed a horizontally-displaced, direct symmetry of the cipher component. Vertically, it shows no symmetry, or if there is symmetry, it is not visible.[2] But when Table I–B is more carefully examined, an invisible, or indirect, vertical symmetry may be discerned where at first glance it is not apparent. If one takes any two *columns* of the table, it is found that the interval between the members of any pair of letters in one column is the same as the interval between the members of the homologous pair of letters in the other column, *if the distance is measured on the cipher component*. For example, consider the 2d and 15th columns (headed by L and I, respectively); take the letters P and G in the 2d column, and J and W in the 15th column. The distance between P and G on the cipher component is 7 intervals; the distance between J and W on the same component is *also* 7 intervals. This phenomenon implies a kind of hidden, or latent, or indirect symmetry within the cipher square. In fact, it may be stated that every table which sets forth in systematic fashion the various secondary alphabets derivable by sliding two primary sequences through all points of coincidence to find cipher equivalents must show some kind of symmetry, both horizontally and

[2] It is true that the first column within the table shows the plain-component sequence, but this is merely because the method of finding the equivalents in this case is such that this sequence is bound to appear in that column, since the successive key letters are A, B, C, . . . Z, and this sequence happens to be identical with the plain component in this case. The same is true of Tables V and XI; it is also applicable to the first row of Tables IX and X.

vertically. The symmetry may be termed *visible* or *direct*, if the sequences of letters in the rows (or columns) are the same throughout and are identical with that of one of the primary components; it may be termed *hidden* or *indirect* if the sequences of letters in the rows or columns are different, apparently not related to either of the components, but are in reality decimations of one of the primary components.

j. When the twelve tables of Appendix 1 are examined in the light of the foregoing remarks, the type of symmetry found in each may be summarized in the following manner:

Table	Horizontal				Vertical			
	Visible or direct		Invisible or indirect		Visible or direct		Invisible or indirect	
	Follows plain component	Follows cipher component	Follows plain component	Follows cipher component	Follows plain component	Follows cipher component	Follows plain component	Follows cipher component
I		x						x
II			x				x	
III		x				x		
IV			x		x			
V		x						x
VI			x				x	
VII	x						x	
VIII	x						x	
IX				x				x
X				x				x
XI			x		x			
XII		x				x		

Of these twelve types of cipher squares, corresponding to the twelve different ways of using a pair of sliding primary components to derive secondary alphabets, the ones best known and most often encountered in cryptographic studies are Tables I–B and II, referred to as being of the Vigenère type; Tables V and VI, referred to as being of the Beaufort type; and Tables IX and X, referred to as being of the Delastelle type. It will be noted that the tables of the Delastelle type show no direct or visible symmetry, either horizontally or vertically and because of this are supposed to yield more security than do any of the other types of tables. But it will presently be shown that the supposed increase in security is more illusory than real.

k. The foregoing facts concerning the various types of quadricular tables generated by diverse methods of using sliding primary components or their equivalent rotating cipher disks will be employed to good advantage, when the studies presently to be undertaken will bring the student to the place where he can comprehend them in the analysis of polyalphabetic systems. But in order not to confuse him with a multiplicity of details which have no direct bearing upon basic principles, one and only one standard method of finding equivalents by means of sliding components will be selected from among the twelve available, as set forth in the preceding subparagraphs. Unless otherwise stated, this method will be the one denoted by the first of the formulae listed in subpar. *f, viz:*

$$\Theta_{k/2}=\Theta_{1/1};\ \Theta_{p/1}=\Theta_{c/2}$$

Calling the plain component "1" and the cipher component "2", this will mean that the keyletter on the cipher component will be set opposite the index, which will be the first letter of the plain component; the plain-text letter to be enciphered will then be sought on the plain component and its equivalent will be the letter opposite it on the cipher component.

SECTION III

THEORY OF SOLUTION OF REPEATING-KEY SYSTEMS

	Paragraph
The three steps in the analysis of repeating-key systems	8
First step: finding the length of the period	9
General remarks on factoring	10
Second step: distributing the cipher text into the component monoalphabets	11
Third step: solving the monoalphabetic distributions	12

8. The three steps in the analysis of repeating-key systems.—*a.* The method of enciphering according to the principle of the repeating key, or repeating alphabets is adequately explained in Section XI of *Elementary Military Cryptography*, and no further reference need be made at this time. The analysis of a cryptogram of this type, regardless of the kind of cipher alphabets employed, or their method of production, resolves itself into three distinct and successive steps.

(1) Determination of the length of the repeating key, which is the same as the determination of the exact number of alphabets involved in the cryptogram;

(2) Allocation or distribution of the letters of the cipher text into the respective cipher alphabets to which they belong. This is the step which reduces the polyalphabetic text to monoalphabetic terms;

(3) Analysis of the individual monoalphabetic distributions to determine plain-text values of the cipher letters in each distribution or alphabet.

b. The foregoing steps will be treated in the order in which mentioned. The first step may be described briefly as that of *determining the period.* The second step may be described briefly as that of *reduction to monoalphabetic terms.* The third step may be designated as *identification of cipher-text values.*

9. First step: finding the length of the period.—*a.* The determination of the period, that is, the length of the key or the number of cipher alphabets involved in a cryptogram enciphered by the repeating-key method is, as a rule, a relatively simple matter. The cryptogram itself usually manifests externally certain phenomena which are the direct result of the use of a repeating key. The principles involved are, however, so fundamental in cryptanalysis that their elucidation warrants a somewhat detailed treatment. This will be done in connection with a short example of encipherment, shown in Fig. 1.

MESSAGE

THE ARTILLERY BATTALION MARCHING IN THE REAR OF THE ADVANCE GUARD KEEPS ITS COMBAT TRAIN WITH IT INSOFAR AS PRACTICABLE.

[Key: BLUE, using direct standard alphabets]

CIPHER ALPHABETS

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
Cipher	(1)	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
	(2)	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
	(3)	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	(4)	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

B L U E	B L U E	B L U E	B L U E
T H E A	A R D K	T H E A	A R D K
		U S Y E	B C X O
R T I L	E E P S	R T I L	E E P S
		S E C P	F P J W
L E R Y	I T S C	L E R Y	I T S C
		M P L C	J E M G
B A T T	O M B A	B A T T	O M B A
		C L N X	P X V E
A L I O	T T R A	A L I O	T T R A
		B W C S	U E L E
N M A R	I N W I	N M A R	I N W I
		O X U V	J Y Q M
C H I N	T H I T	C H I N	T H I T
		D S C R	U S C X
G I N T	I N S O	G I N T	I N S O
		H T H X	J Y M S
H E R E	F A R A	H E R E	F A R A
		I P L I	G L L E
A R O F	S P R A	A R O F	S P R A
		B C I J	T A L E
T H E A	C T I C	T H E A	C T I C
		U S Y E	D E C G
D V A N	A B L E	D V A N	A B L E
		E G U R	B M F I
C E G U		C E G U	
		D P A Y	
a	*a*	*b*	*b*

CRYPTOGRAM

U S Y E S E C P M P L C C L N X B W C S O X U V D S C R H T

H X I P L I B C I J U S Y E E G U R D P A Y B C X O F P J W

J E M G P X V E U E L E J Y Q M U S C X J Y M S G L L E T A

L E D E C G B M F I

FIGURE 1.

b. Regardless of what system is used, identical plain-text letters enciphered by the same cipher alphabet [1] must yield identical cipher letters. Referring to Fig. 1, such a condition is brought about every time that identical plain-text letters happen to be enciphered with the same key-letter, or every time identical plain-text letters fall into the same column in the encipherment.[2] Now since the number of columns or positions with respect to the key is very limited (except in the case of very *long* key words), and since the repetition of letters is an inevitable condition in plain text, it follows that there will be in a message of fair length many cases where identical plain-text letters *must* fall into the same column. They will thus be enciphered by the same cipher alphabet, resulting, therefore, in the production of many identical letters in the cipher text and these will represent identical letters in the plain text. When identical plain-text polygraphs fall into identical columns the result is the formation of identical cipher-text polygraphs, that is, repetitions of groups of 2, 3, 4, . . . letters are exhibited in the cryptogram. Repetitions of this type will hereafter be called *causal repetitions*, because they are produced by a definite, traceable cause, *viz*, the encipherment of identical letters by the same cipher alphabets.

c. It will also happen, however, that *different* plain-text letters falling in *different* columns will, by mere accident, produce identical cipher letters. Note, for example, in Fig. 1 that in Column 1, R_p becomes S_c and that in Column 2, H_p also becomes S_c. The production of an identical cipher text letter in these two cases (that is, a repetition where the plain-text letters are different and enciphered by different alphabets) is merely fortuitous. It is, in every day language, "a mere coincidence", or "an accident." For this reason repetitions of this type will hereafter be called *accidental repetitions*.

d. A consideration of the phenomenon pointed out in *c* makes it obvious that in polyalphabetic ciphers it is important that the cryptanalyst be able to tell whether the repetitions he finds in a specific case are causal or accidental in their origin, that is, whether they represent actual enciphermentsof identical plain-text letters by identical keying elements, or mere coincidences brought about purely fortuitously.

e. Now accidental repetitions will, of course, happen fairly frequently with individual letters, but less frequently with digraphs, because in this case the same kind of an "accident" must take place twice in succession. Intuitively one feels that the chances that such a purely fortuitous coincidence will happen two times in succession must be much less than that it will happen every once in a while in the case of single letters. Similarly, intuition makes one feel that the chances of such accidents happening in the case of three or more consecutive letters are still less than in the case of digraphs, decreasing very rapidly as the repetition increases in length.

f. The phenomena of cryptographic repetition may, fortunately, be dealt with statistically, thus taking the matter outside the realm of intuition and putting it on a firm mathematical or objective basis. Moreover, often the statistical analysis will tell the cryptanalyst when he has arranged or rearranged his text properly, that is, when he is approaching or has reached monoalphabeticity in his efforts to reduce polyalphabetic text to its simplest terms. However, in order to preserve continuity of thought it is deemed inadvisable to inject these statistical considerations at this place in the text proper; they have been incorporated in Appendix 2 hereof. The student is advised to study the Appendix very carefully after he has finished reading this section of the text.

g. At this point it will merely be indicated that if a cryptanalyst were to have at hand only the cryptogram of Fig. 1, with the repetitions underlined as below, a statistical study of the

[1] It is to be understood, of course, that cipher alphabets with single equivalents are meant in this case.

[2] The frequency with which this condition may be *expected* to occur can be definitely calculated. A discussion of this point falls beyond the scope of the present text.

number and length of the repetitions within the message (Par. 5 of Appendix 2) would tell him that while some of the digraphic repetitions may be accidental, the chances that they all are accidental are small. In the case of the tetragraphic repetition he would realize that the chances of its being accidental are very small indeed.

A.	U S Y E S	E C P M P L C C L N	X B W C S	O X U V D
B.	S C R H T	H X I P L I B C I J	U S Y E E	G U R D P
C.	A Y B C X	O F P J W J E M G P	X V E U E	L E J Y Q
D.	M U S C X	J Y M S G L L E T A	L E D E C	G B M F I

h. A consideration of the facts therefore leads to but one conclusion, *viz*, that the repetitions exhibited by the cryptogram under investigation are *not accidental* but are *causal* in their origin; and the cause is in this case not difficult to find: repetitions in the plain text were actually enciphered by identical alphabets. In order for this to occur, it was necessary that the tetragraph USYE, for example, fall *both* times in *exactly* the same relative position with respect to the key. Note, for example, that UYSE in Fig. 1 represents in both cases the plain-text polygraph THEA. The first time it occurred it fell in positions 1–2–3–4 with respect to the key; the second time it occurred it happened to fall in the very same relative positions, although it might just as well have happened to fall in any of the other three possible relative positions with respect to the key, *viz*, 2–3–4–1, 3–4–1–2, or 4–1–2–3.

i. Lest the student be misled, however, a few more words are necessary on this subject. In the preceding subparagraph the word "happened" was used; this word correctly expresses the idea in mind, because the insertion or deletion of a single plain-text letter between the two occurrences would have thrown the second occurrence one letter forward or backward, respectively, and thus caused the polygraph to be enciphered by a sequence of alphabets such as can no longer produce the cipher polygraph USYE from the plain-text polygraph THEA. On the other hand, the insertion or deletion of this one letter might bring the letters of some other polygraph into similar columns so that some other repetition would be exhibited in case the USYE repetition had thus been suppressed.

j. The encipherment of similar letters by similar cipher alphabets is therefore the *cause* of the production of repetitions in the cipher text in the case of repeating-key ciphers. What principles can be derived from this fact, and how can they be employed in the solution of cryptograms of this type?

k. If a count is made of the number of letters from and including the first USYE to, but not including, the second occurrence of USYE, a total of 40 letters is found to intervene between the two occurrences. This number, 40, must, of course, be an exact multiple of the length of the key. Having the plain-text before one, it is easily seen that it is the 10th multiple; that is, the 4-letter key has repeated itself 10 times between the first and the second occurrence of USYE. It follows, therefore, that if the length of the key were not known, the number 40 could safely be taken to be an exact multiple of the length of the key; in other words, one of the *factors* of the number 40 would be equal to the length of the key. The word "safely" is used in the preceding sentence to mean that the interval 40 applies to a repetition of 4 letters and it has been shown that the chances that this repetition is accidental are small. The factors of 40 are 2, 4, 5, 8, 10, and 20. So far as this single repetition of USYE is concerned, if the length of the key were not known, all that could be said about the latter would be that it is equal to one of these factors. The repetition by itself gives no further indications. How can the exact factor be selected from among a list of several possible factors?

152018—38——2

l. Let the intervals between all the repetitions in the cryptogram be listed. They are as follows:

Repetition	Interval	Factors
1st USYE to 2d USYE	40	2, 4, 5, 8, 10, 20.
1st BC to 2d BC	16	2, 4, 8.
1st CX to 2d CX	25	5.
1st EC to 2d EC	88	2, 4, 11, 22, 44.
1st LE to 2d LE	16	2, 4, 8.
2d LE to 3d LE	4	2, 4.
1st LE to 3d LE	20	2, 4, 5, 10.
1st JY to 2d JY	8	2, 4.
1st PL to 2d PL	24	2, 3, 4, 6, 8, 10, 12.
1st SC to 2d SC	52	2, 4, 13, 26.
(1st SY to 2d SY, already included in USYE.)		
(1st US to 2d US, already included in USYE.)		
2d US to 3d US	36	2, 3, 4, 6, 9, 18.
(1st US to 3d US, already included in USYE.)		
(1st YE to 2d YE, already included in USYE.)		

m. Are all these repetitions *causal* repetitions? It can be shown (Appendix 2, par. 4*c*) that the odds against a theory that the UYSE repetition is accidental are about 99 to 1 (since the probability for its occurrence is .01). It can also be shown that the odds against a theory that the 10 digraphs which occur two or more times are accidental repetitions are over 4 to 1 (Appendix 2, par. 5*c*); the odds against a theory that the two digraphs which occur 3 times are accidental repetitions are quite large. (Probability is calculated to be about .06.) The chances are very great, therefore, that all or nearly all these repetitions are causal. Certainly the chances against the two occurrences of the tetragraph UYSE and the three occurrences of the two different digraphs (LE and US) being accidental are quite high, and it is therefore not astonishing that the intervals between all the various repetitions, except in one case, contain the factors 2 and 4.

n. This means that if the cipher is written out in either 2 columns or 4 columns, all these repetitions (except the CX repetition) would fall into the same columns. From this it follows that the length of the key is either 2 or 4, the latter, on practical grounds, being more probable than the former. Doubts concerning the matter of choosing between a 2-letter and a 4-letter key will be dissolved when the cipher text is distributed into its component uniliteral frequency distributions.

o. The repeated digraph CX in the foregoing message is an accidental repetition, as will be apparent by referring to Fig. 1. Had the message been longer there would have been more such accidental repetitions, but, on the other hand, there would be a proportionately greater number of causal repetitions. This is because the phenomenon of repetition in plain text is so all-pervading.

p. Sometimes it happens that the cryptanalyst quickly notes a repetition of a polygraph of four or more letters, the interval between the first and second occurrences of which has only two factors, of which one is a relatively small number, the other a relatively high incommensurable number. He may therefore assume at once that the length of the key is equal to the smaller factor without searching for additional recurrences upon which to corroborate his assumption. Suppose, for example, that in a relatively short cryptogram the interval between the first and second occurrences of a polygraph of five letters happens to be a number such as 203, the factors of which are 7 and 29. Evidently the number of alphabets may at once be

assumed to be 7, unless one is dealing with messages exchanged among correspondents known to use long keys. In the latter case one could assume the number of alphabets to be 29.

g. The foregoing method of determining the period in a polyalphabetic cipher is commonly referred to the literature as "factoring the intervals between repetitions"; or more often it is simply called "factoring." Because the latter is an apt term and is brief, it will be employed hereafter in this text to designate the process.

10. General remarks on factoring.—*a.* The statement made in Par. 2 with respect to the cyclic phenomena said to be exhibited in cryptograms of the periodic type now becomes clear. The use of a short repeating key produces a periodicity of recurrences or repetitions collectively termed "cyclic phenomena", an analysis of which leads to a determination of the length of the period or cycle, and this gives the length of the key. Only in the case of relatively short cryptograms enciphered by a relatively long key does factoring fail to lead to the correct determination of the number of cipher alphabets in a repeating-key cipher; and of course, the fact that a cryptogram contains repetitions whose factors show constancy is in itself an indication and test of its periodic nature. It also follows that if the cryptogram is not a repeating-key cipher, then factoring will show no definite results, and conversely the fact that it does not yield definite results at once indicates that the cryptogram is not a periodic, repeating-key cipher.

b. There are two cases in which factoring leads to no definite results. One is in the case of monoalphabetic substitution ciphers. Here recurrences are very plentiful as a rule, and the intervals separating these recurrences may be factored, *but the factors will show no constancy;* there will be several factors common to many or most of the recurrences. This in itself is an indication of a monoalphabetic substitution cipher, if the very fact of the presence of many recurrences fails to impress itself upon the inexperienced cryptanalyst. The other case in which the process of factoring is nonsignificant involves certain types of nonperiodic, polyalphabetic ciphers. In certain of these ciphers recurrences of digraphs, trigraphs, and even polygraphs may be plentiful in a long message, but the intervals between such recurrences bear no definite multiple relation to the length of the key, such as in the case of the true periodic, repeating-key cipher, in which the alphabets change with successive letters and repeat themselves over and over again.

c. Factoring is not the only method of determining the length of the period of a periodic, polyalphabetic substitution cipher, although it is by far the most common and easily applied. At this point it will merely be stated that when the message under study is relatively short in comparison with the length of the key, so that there are only a few cycles of cipher text and no long repetitions affording a basis for factoring, there are several other methods available. However, it being deemed inadvisable to interject the data concerning those other methods at this point, they will be explained subsequently. It is desirable at this juncture merely to indicate that methods other than factoring do exist and are used in practical work.

d. Fundamentally, the factoring process is merely a more or less simple mathematical method of studying the phenomena of periodicity in cryptograms. It will usually enable the cryptanalyst to ascertain definitely whether or not a given cryptogram is periodic in nature, and if so, the length of the period, *stated in terms of the cryptographic unit involved.* By the latter statement is meant that the factoring process may be applied not only in analyzing the periodicity manifested by cryptograms in which the plain-text units subjected to cryptographic treatment are monographic in nature (i. e. are single letters) but also in studying the periodicity exhibited by those occasional cryptograms wherein the plain-text units are digraphic, trigraphic, or n-graphic in character. The student should bear this point in mind when he comes to the study of substitution systems of the latter sort. However, the present text will deal solely with cases of the former type, wherein the plain-text units subjected to cryptographic treatment are single letters.

11. Second step: distributing the cipher text into the component monoalphabets.—*a.* After the number of cipher alphabets involved in the cryptogram has been ascertained, the next step is to rewrite the message in groups corresponding to the length of the key, or in columnar fashion, whichever is more convenient, and this automatically divides up the text so that the letters belonging to the same cipher alphabet occupy similar positions in the groups, or, if the columnar method is used, fall in the same column. The letters are thus allocated or distributed into the respective cipher alphabets to which they belong. This reduces the polyalphabetic text to monoalphabetic terms.

b. Then separate uniliteral frequency distributions for the thus isolated individual alphabets are compiled. For example, in the case of the cipher on page 13, having determined that four alphabets are involved, and having rewritten the message in four columns, a frequency distribution is made of the letters in Column 1, another is made of the letters in Column 2, and so on for the rest of the columns. *Each of the resulting distributions is therefore a monoalphabetic frequency distribution.* If these distributions do not give the characteristic irregular crest and trough appearance of monoalphabetic frequency distributions, then the analysis which led to the hypothesis as regards the number of alphabets involved is fallacious. In fact, the appearance of these individual distributions may be considered to be an index of the correctness of the factoring process; for theoretically, and practically, the individual distributions constructed upon the *correct* hypothesis will tend to conform more closely to the irregular crest and trough appearacne of a monoalphabetic frequency distribution than will the graphic tables constructed upon an incorrect hypothesis. These individual distributions may also be tested for monoalphabeticity by statistical methods.

12. Third step: solving the monoalphabetic distributions.—The difficulty experienced in analyzing the individual or isolated frequency distributions depends mostly upon the type of cipher alphabets that is used. It is apparent that mixed alphabets may be used just as easily as standard alphabets, and, of course, the cipher letters themselves give no indication as to which is the case. However, just as it was found that in the case of monoalphabetic substitution ciphers, a uniliteral frequency distribution gives clear indications as to whether the cipher alphabet is a standard or a mixed alphabet, by the relative positions and extensions of the crests and troughs in the table, so it is found that in the case of repeating-key ciphers, uniliteral frequency distributions for the isolated or individual alphabets will also give clear indications as to whether these alphabets are standard alphabets or mixed alphabets. Only one or two such frequency distributions are necessary for this determination; if they appear to be standard alphabets, similar distributions can be made for the rest of the alphabets; but if they appear to be mixed alphabets, then it is best to compile triliteral frequency distributions for all the alphabets. The analysis of the values of the cipher letters in each table proceeds along the same lines as in the case of monoalphabetic ciphers. The analysis is more difficult only because of the reduced size of the tables, but if the message be very long, then each frequency distribution will contain a sufficient number of elements to enable a speedy solution to be achieved.

SECTION IV

REPEATING-KEY SYSTEMS WITH STANDARD CIPHER ALPHABETS

	Paragraph
Solution by applying principles of frequency	13
Solution by completing the plain-component sequence	14
Solution by the "probable-word method"	15

13. Solution by applying principles of frequency.—*a.* In the light of the foregoing principles, let the following cryptogram be studied:

MESSAGE

	1	2	3	4	5
A.	A U K H Y	J A M K I	Z Y M W M	J M I G X	N F M L X
B.	E T I M I	Z H B H R	A Y M Z M	I L V M E	J K U T G
C.	D P V X K	Q U K H Q	L H V R M	J A Z N G	G Z V X E
D.	N L U F M	P Z J N V	C H U A S	H K Q G K	I P L W P
E.	A J Z X I	G U M T V	D P T E J	E C M Y S	Q Y B A V
F.	A L A H Y	P O E X W	P V N Y E	E Y X E E	U D P X R
G.	B V Z V I	Z I I V O	S P T E G	K U B B R	Q L L X P
H.	W F Q G K	N L L L E	P T I K W	D J Z X I	G O I O I
J.	Z L A M V	K F M W F	N P L Z I	O V V F M	Z K T X G
K.	N L M D F	A A E X I	J L U F M	P Z J N V	C A I G I
L.	U A W P R	N V I W E	J K Z A S	Z L A F M	H S

A search for repetitions discloses the following short list with the intervals and factors above 10 omitted (for previous experience may lead to the conclusion that it is unlikely that the cryptogram involves more than 10 alphabets, showing the number of recurrences which it does):

Repetition	Location	Interval	Factors
LUFMPZJNVC	D1, K3	160	2, 4, 5, 8, 10.
JZXIG	E1, H4	90	2, 3, 5, 6, 9, 10.
EJK	B4, L2	215	5.
PTE	E3, G3	50	2, 5, 10.
QGK	D4, H1	85	5.
UKH	A1, C2	55	5.
ZLA	J1, L4	65	5.
AS	D3, L3	175	3, 5, 7,
EJ	B4, L2	115	5.
FM	A5, D1	57	3.
FM	A5, J2	185	5.
FM	J2, J4	12	2, 3, 4, 6.
FM	J4, K3	20	2, 4, 5, 10.
FM	K3, L4	30	2, 3, 5, 6, 10.
JA	A2, C4	60	2, 3, 4, 5, 6, 10.
LA	F1, J1	75	3, 5.
LA	J1, L4	65	5.
LL	G5, H2	10	2, 5.
NL	D1, H2	105	3, 5, 7.
NL	H2, K1	45	3, 5, 9.
VX	C1, C5	20	2, 4, 5, 10.
YM	A3, B3	25	5.

b. The factor 5 appears in all but two cases, each of which involves only a digraph. It seems almost certain that the number of alphabets is five. Since the text already appears in groups of five letters, it is unnecessary to rewrite the message. The next step is to make a uniliteral frequency distribution for Alphabet 1 to see if it can be determined whether or not standard alphabets are involved. It is as follows:

ALPHABET 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
5	1	2	3	4	5	2	3	2	5	2	1	5	5	1	5	2	3	1	2	3	1	2	1	1	5

c. Although the indications are not very clear cut, yet if one takes into consideration the small amount of data the assumption of a direct standard alphabet with $W_c=A_p$, is worth further test. Accordingly a similar distribution is made for Alphabet 2.

ALPHABET 2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
5	1	1	1	3	3	2	3	1	2	4	5	1	2	2	5	2	1	1	2	4	4	3	2	4	2

d. There is every indication of a direct standard alphabet, with $H_c=A_p$. Let similar distributions be made for the last three alphabets. They are as follows:

ALPHABET 3

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	2	1	1	2	1	1	1	5	2	2	3	5	1	1	2	2	1	3	3	4	5	1	1	1	5

ALPHABET 4

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3	1	1	1	3	3	4	4	1	1	2	2	3	4	1	1	1	1	2	2	2	4	5	5	2	2

ALPHABET 5

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	1	1	1	5	2	3	5	5	1	2	1	5	1	1	2	1	4	3	1	1	5	2	2	2	1

e. After but little experiment it is found that the distributions can best be made to fit the normal when the following values are assumed:

Alphabet 1..... $A_p=W_c$
Alphabet 2..... $A_p=H_c$
Alphabet 3..... $A_p=I_c$
Alphabet 4..... $A_p=T_c$
Alphabet 5..... $A_p=E_c$

f. Note the key word given by the successive equivalents of A_p: WHITE. The real proof of the correctness of the analysis is, of course, to test the values of the solved alphabets on the cryptogram. The five complete cipher alphabets are as follows:

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
Cipher	1	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	2	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
	3	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
	4	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	5	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D

FIGURE 2.

g. Applying these values to the first few groups of our message, the following is found:

	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
Cipher.....	A	U	K	H	Y	J	A	M	K	I	Z	Y	M	W	M	J	M	I	G	X	N	F	M	L	X	. . .
Plain.....	E	N	C	O	U	N	T	E	R	E	D	R	E	D	I	N	F	A	N	T	R	Y	E	S	T	. . .

h. Intelligible text at once results, and the solution can now be completed very quickly. The complete message is as follows:

ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MACHINE GUN COMPANY IN TRUCKS NEAR EMMITSBURG. AM HOLDING MIDDLE CREEK NEAR HILL 543 SOUTHWEST OF FAIRPLAY. WHEN FORCED BACK WILL CONTINUE DELAYING REDS AT MARSH CREEK. HAVE DESTROYED BRIDGES ON MIDDLE CREEK BETWEEN EMMITSBURG–TANEYTOWN ROAD AND RHODES MILL.

i. In the foregoing example (which is typical of the system erroneously attributed, in cryptographic literature, to the French cryptographer Vigenère, although to do him justice, he made no claim of having "invented" it), direct standard alphabets were used, but it is obvious that reversed standard alphabets may be used and the solution accomplished in the same manner. In fact, the now obsolete cipher disk used by the United States Army for a number of years yields exactly this type of cipher, which is also known in the literature as the Beaufort Cipher, and by other names. In fitting the isolated frequency distributions to the normal, the direction of "reading" the crests and troughs is merely reversed.

14. Solution by completing the plain-component sequence.—*a.* There is another method of solving this type of cipher, which is worthwhile explaining, because the underlying principles will be found useful in many cases. It is a modification of the method of solution by completing the plain-component sequence, already explained in *Military Cryptanalysis*, Part I.

b. After all, the individual alphabets of a cipher such as the one just solved are merely direct standard alphabets. It has been seen that monoalphabetic ciphers in which standard cipher alphabets are employed may be solved almost mechanically by completing the plain-component sequence. The plain text reappears on only one generatrix and this generatrix is the same for the whole message. It is easy to pick this generatrix out of all the other generatrices because it is the only one which yields intelligible text. Is it not apparent that if the same process is applied to the cipher letters of the *individual alphabets* of the cipher just solved that the plain-text equivalents of these letters must all reappear on one and the same generatrix? But how will the generatrix which actually contains the plain-text letters be distinguishable from the other generatrices, since these plain-text letters are not consecutive letters in the plain text but only letters separated from one another by a constant interval? The answer is simple. The plain-text generatrix should be distinguishable from the others *because it will show more and a better assortment of high-frequency letters, and can thus be selected by the eye from the whole set of generatrices.* If this is done with all the alphabets in the cryptogram, it will merely be necessary to assemble the letters of the thus selected generatrices in proper order, and the result sould be consecutive letters forming intelligible text.

c. An example will serve to make the process clear. Let the same message be used as before. Factoring showed that it involves five alphabets. Let the first ten cipher letters *in each alphabet* be set down in a horizontal line and let the normal alphabet sequences be completed. Thus:

	Alphabet 1	Alphabet 2	Alphabet 3	Alphabet 4	Alphabet 5
1	AJZJNEZAIJ	UAYMFTHYLK	KMMIMIBMVU	HKWGLMHZMT	YIMXXIRMEG
2	BKAKOFABJK	VBZNGUIZML	LNNJNJCNWV	ILXHMNIANU	ZJNYYJSNFH
3	CLBLPGBCKL	WCAOHVJANM	MOOKOKDOXW	JMYINOJBOV	AKOZZKTOGI
4	DMCMQHCDLM	XDBPIWKBON	NPPLPLEPYX	KNZJOPKCPW	BLPAALUPHJ
5	ENDNRIDEMN	YECQJXLCPO	OQQMQMFQZY	LOAKPQLDQX	CMQBBMVQIK
6	FOEOSJEFNO	ZFDRKYMDQP	PRRNRNGRAZ	MPBLQRMERY	DNRCCNWRJL
7	GPFPTKFGOP	AGESLZNERQ	QSSOSOHSBA	NQCMRSNFSZ	EOSDDOXSKM
8	HQGQULGHPQ	BHFTMAOFSR	RTTPTPITCB	ORDNSTOGTA	FPTEEPYTLN
9	IRHRVMHIQR	CIGUNBPGTS	SUUQUQJUDC	PSEOTUPHUB	GQUFFQZUMO
10	JSISWNIJRS	DJHVOCQHUT	TVVRVRKVED	QTFPUVQIVC	HRVGGRAVNP
11	KTJTXOJKST	EKIWPDRIVU	UWWSWSLWFE	RUGQVWRJWD	ISWHHSBWOQ
12	LUKUYPKLTU	FLJXQESJWV	VXXTXTMXGF	SVHRWXSKXE	JTXIITCXPR
13	MVLVZQLMUV	GMKYRFTKXW	WYYUYUNYHG	TWISXYTLYF	KUYJJUDYQS
14	NWMWARMNVW	HNLZSGULYX	XZZVZVOZIH	UXJTYZUMZG	LVZKKVEZRT
15	OXNXBSNOWX	IOMATHVMZY	YAAWAWPAJI	VYKUZAVNAH	MWALLWFASU
16	PYOYCTOPXY	JPNBUIWNAZ	ZBBXBXQBKJ	WZLVABWOBI	NXBMMXGBTV
17	QZPZDUPQYZ	KQOCVJXOBA	ACCYCYRCLK	XAMWBCXPCJ	OYCNNYHCUW
18	RAQAEVQRZA	LRPDWKYPCB	BDDZDZSDML	YBNXCDYQDK	PZDOOZIDVX
19	SBRBFWRSAB	MSQEXLZQDC	CEEAEATENM	ZCOYDEZREL	QAEPPAJEWY
20	TCSCGXSTBC	NTRFYMARED	DFFBFBUFON	ADPZEFASFM	RBFQQBKFXZ
21	UDTDHYTUCD	OUSGZNBSFE	EGGCGCVGPO	BEQAFGBTGN	SCGRRCLGYA
22	VEUEIZUVDE	PVTHAOCTGF	FHHDHDWHQP	CFRCGHCUHO	TDHSSDMHZB
23	WFVFJAVWEF	QWUIBPDUHG	GIIEIEXIRQ	DGSCHIDVIP	UEITTENIAC
24	XGWGKBWXFG	RXVJCQEVIH	HJJFJFYJSR	EHTDIJEWJQ	VFJUUFOJBD
25	YHXHLCXYGH	SYWKDRFWJI	IKKGKGZKTS	FIUEJKFXKR	WGKVVGPKCE
26	ZIYIMDYZHI	TZXLESGXKJ	JLLHLHALUT	GJVFKLGYLS	XHLWWHQLDF

Figure 3.

d. If the high-frequency generatrices underlined in Figure 3 are selected and their letters are juxtaposed *in columns* the consecutive letters of intelligible plain text immediately present themselves. Thus:

Selected Generatrices:

For Alphabet 1, generatrix 5	E	N	D	N	R	I	D	E	M	N
For Alphabet 2, generatrix 20	N	T	R	F	Y	M	A	R	E	D
For Alphabet 3, generatrix 19	C	E	E	A	E	A	T	E	N	M
For Alphabet 4, generatrix 8	O	R	D	N	S	T	O	G	T	A
For Alphabet 5, generatrix 23	U	E	I	T	T	E	N	I	A	C

Columnar juxtaposition of letters from selected generatrices:

1	2	3	4	5
E	N	C	O	U
N	T	E	R	E
D	R	E	D	I
N	F	A	N	T
R	Y	E	S	T
I	M	A	T	E
D	A	T	O	N
E	R	E	G	I
M	E	N	T	A
N	D	M	A	C

Figure 4.

Plain text: ENCOUNTERED RED INFANTRY ESTIMATED AT ONE REGIMENT AND MAC . . . .

e. Solution by this method can thus be achieved without the compilation of any frequency tables whatever and is very quickly attained. The inexperienced cryptanalyst may have difficulty at first in selecting the generatrices which contain the most and the best assortment of high-frequency letters, but with increased practice, a high degree of proficiency is attained. After all it is only a matter of experiment, trial, and error to select and assemble the proper generatrices so as to produce intelligible text.

f. If the letters on the sliding strips were accompanied by numbers representing their relative frequencies in plain text, and these numbers were added *across* each generatrix, then that generatrix with the highest total frequency would *theoretically* always be the plain-text generatrix. Practically it will be among the generatrices which show the first three or four greatest totals. Thus, an entirely mathematical solution for this type of cipher may be applied.

g. If the cipher alphabets are reversed standard alphabets, it is only necessary to convert the cipher letters of each isolated alphabet into their normal, plain-component equivalents and then proceed as in the case of direct standard alphabets.

h. It has been seen how the key word may be discovered in this type of cryptogram. Usually the key is made up of those letters in the successive alphabets whose equivalents are A_p, but other conventions are of course possible. Sometimes a key number is used, such as 8–4–7–1–12, which means merely that A_p is represented by the eighth letter from A (in the normal alphabet) in the first cipher alphabet, by the fourth letter from A in the second cipher alphabet, and so on. This modification is known in the literature as the Gronsfeld cipher. However, the method of solution as illustrated above, being independent of the nature of the key, is the same as before.

15. Solution by the "probable-word method."—*a.* The common use of key words in cryptograms such as the foregoing makes possible a method of solution that is simple and can be used where the more detailed method of analysis using frequency distributions or by completing the plain-component sequence is of no avail. In the case of a very short message which may show no recurrences and give no indications as to the number of alphabets involved, this modified method will be found most useful.

b. Briefly, the method consists in assuming the presence of a probable word in the message, and referring to the alphabets to find the key letters applicable when this hypothetical word is assumed to be present in various positions in the cipher text. If the assumed word happens to be correct, and is placed in the correct position in the message, the key letters produced by referring to the alphabets will yield the key word. In the following example it is assumed that reversed standard alphabets are known to be used by the enemy.

MESSAGE

M D S T J L Q C X C K Z A S A N Y Y K O L P

c. Extraneous circumstances lead to the assumption of the presence of the word AMMUNITION. One may assume that this word begins the message. Using sliding normal components, one reversed, the other direct, the key letters are ascertained by noting what the successive equivalents of A_p are. Thus:

Cipher	M D S T J L Q C X C
Plain text	A M M U N I T I O N
"Key"	M P E N W T J K L P

The key does not spell any intelligible word. One therefore shifts the assumed word one letter forward and another trial is made.

Cipher	D	S	T	J	L	Q	C	X	C	K
Plain text	A	M	M	U	N	I	T	I	O	N
"Key"	D	E	F	D	Y	Y	V	F	Q	X

This also yields no intelligible key word. One continues to shift the assumed word forward one space at a time until the following point is reached.

Cipher	L	Q	C	X	C	K	Z	A	S	A
Plain text	A	M	M	U	N	I	T	I	O	N
"Key"	L	C	O	R	P	S	S	I	G	N

The key now becomes evident. It is a cyclic permutation of SIGNAL CORPS. It should be clear that since the key word or key phrase repeats itself during the encipherment of such a message, the plain-text word upon whose assumed presence in the message this test is being based may begin to be enciphered at any point in the key, and continue over into its next repetition if it is longer than the key. When this is the case it is merely necessary to shift the latter part of the sequence of key letters to the first part, as in the case noted: LCORPSSIGN is transposed into SIGN . . . LCORPS, and thus SIGNAL CORPS.

d. It will be seen in the foregoing method of solution that the length of the key is of no particular interest or consequence in the steps taken in effecting the solution. The determination of the length and elements of the key comes after the solution rather than before it. In this case the length of the period is seen to be eleven, corresponding to the length of the key (SIGNAL CORPS).

e. The foregoing method is one of the other methods of determining the length of the key (besides factoring), referred to in Par. 10*c*.

f. If the assumption of reversed standard alphabets yields no good results, then direct standard alphabets are assumed and the test made exactly in the same manner. As will be shown subsequently, the method can also be used as a last resort when mixed alphabets are employed.

g. When the assumed word is longer than the key, the sequence of recovered key letters will show a periodicity equal to the length of the key; that is, after a certain number of letters the sequence of key letters will repeat. This phenomenon would be most useful in the case of keys that are not intelligible words but are composed of random letters or figures. Of course, if such a key is longer than the assumed word, this method is of no avail.

h. This method of solution by searching for a word is contingent upon the following circumstances:

(1) That the word whose presence is assumed actually occurs in the message, is properly spelled, and correctly enciphered.

(2) That the sliding components (or equivalent cipher disks or squares) employed in the search for the assumed word are actually the ones which were employed in the encipherment, or are such as to give identical results as the ones which were actually used.

(3) That the pair of enciphering equations used in the test is actually the pair which was employed in the encipherment; or if a cipher square is used in the test, the method of finding equivalents gives results that correspond with those actually obtained in the encipherment. (See par. 9.)

i. The foregoing appears to be quite an array of contingencies and the student may think that on this account the method will often fail. But examining these contingencies one by one, it will be seen that successful application of the method may not be at all rare—after the solution of some messages has disclosed what sort of paraphernalia and methods of employing them are favored by the enemy. From the foregoing remark it is to be inferred that the probable-word method has its greatest usefulness not in an initial solution of a system, but only after successful study of enemy communications by more difficult processes of analysis has told its story to the alert cryptanalyst. Although it is commonly attributed to Bazeries, the French cryptanalyst of 1900, the probable-word method is very old in cryptanalysis and goes back several centuries. Its usefulness in practical work may best be indicated by quoting from a competent observer [1]:

> There is another [method] which is to this first method what the geometric method is to analysis in certain sciences, and, according to the whims of individuals, certain cryptanalysts prefer one to the other. Certain others, incapable of getting the answer with one of the methods in the solution of a difficult problem, conquer it by means of the other, with a disconcerting masterly stroke. This other method is that of the probable word. We may have more or less definite opinions concerning the subject of the cryptogram. We may know something about its date, and the correspondents, who may have been indiscreet in the subject they have treated. On this basis, the hypothesis is made that a certain word probably appears in the text. . . . In certain classes of documents, military or diplomatic telegrams, banking and mining affairs, etc., it is not impossible to make very important assumptions about the presence of certain words in the text. After a cryptanalyst has worked for a long time with the writings of certain correspondents, he gets used to their expressions. He gets a whole load of words to try out; then the changes of key, and sometimes of system, no longer throw into his way the difficulties of an absolutely new study, which might require the analytical method.

[1] Givierge, M., *Cours de Cryptographie*, Paris, 1925, p. 30.

SECTION V

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, I

	Paragraph
Reason for the use of mixed alphabets	16
Interrelated mixed alphabets	17
Principles of direct symmetry of position	18
Initial steps in the solution of a typical example	19
Application of principles of direct symmetry of position	20
Subsequent steps in solution	21
Completing the solution	22
Solution of subsequent messages enciphered by same cipher component	23
Summation of relative frequencies as an aid to the selection of the correct generatrices	24
Solution by the probable-word method	25
Solution when plain component is mixed, the cipher component, the normal	26

16. Reason for the use of mixed alphabets.—*a.* It has been seen in the examples considered thus far that the use of several alphabets in the same message does not greatly complicate the analysis of such a cryptogram. There are three reasons why this is so. Firstly, only relatively few alphabets were employed; secondly, these alphabets were employed in a periodic or repeating manner, giving rise to cyclic phenomena in the cryptogram, by means of which the number of alphabets could be determined; and, thirdly, the cipher alphabets were *known* alphabets, by which is meant merely that the sequences of letters in both components of the cipher alphabets were known sequences.

b. In the case of monoalphabetic ciphers it was found that the use of a mixed alphabet delayed the solution to a considerable degree, and it will now be seen that the use of mixed alphabets in polyalphabetic ciphers renders the analysis much more difficult than the use of standard alphabets, but the solution is still fairly easy to achieve.

17. Interrelated mixed alphabets.—*a.* It was stated in Par. 5 that the method of producing the mixed alphabets in a polyalphabetic cipher often affords clues which are of great assistance in the analysis of the cipher alphabets. This is so, of course, only when the cipher alphabets are interrelated secondary alphabets produced by sliding components or their equivalents. Reference is now made to the classification set forth in Par. 6, in connection with the types of alphabets which may be employed in polyalphabetic substitution. It will be seen that thus far only Cases A (1) and (2) have been treated. Case B (1) will now be discussed.

b. Here one of the components, the plain component, is the normal sequence, while the cipher component is a mixed sequence, the various juxtapositions of the two components yielding mixed alphabets. The mixed component may be a systematically-mixed or a random-mixed sequence. If the 25 successive displacements of the mixed component are recorded in separate lines, a symmetrical cipher square such as that shown in Fig. 5 results therefrom. It is identical in form with the square table shown on p. 7, labeled Table I–A.

Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L
	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E
	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A
	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V
	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N
	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W
	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O
	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R
	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T
	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B
Cipher.....	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C
	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D
	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F
	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G
	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I
	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J
	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K
	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M
	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P
	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q
	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U
	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X
	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y

FIGURE 6.

c. Such a cipher square may be used in exactly the same manner as the Vigenère square. With the key word BLUE and conforming to the normal enciphering equations ($\Theta_{k/2}=\Theta_{l/1}$; $\Theta_{y/1}=\Theta_{e/2}$), the following lines of the square would be used:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L

FIGURE 6*a*.

These lines would, of course, yield the following cipher alphabets:

(1)	Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher.....	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H
(2)	Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher.....	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z
(3)	Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher.....	U	X	Y	Z	L	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S
(4)	Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher.....	E	A	V	N	W	O	R	T	H	B	C	D	F	G	I	J	K	M	P	Q	S	U	X	Y	Z	L

FIGURE 6*b*.

18. Principles of direct symmetry of position.—*a.* It was stated directly above that Fig. 5 is a symmetrical cipher square, by which is meant that the letters in its successive horizontal lines show a *symmetry of position* with respect to one another. They constitute, in reality, one and only one sequence or series of letters, the sequences being merely displaced successively 1, 2, 3, . . . intervals. The symmetry exhibited is obvious and is said to be visible, or *direct*. This fact can be used to good advantage, as has already been alluded to in par. 7*j*.

b. Consider, for example, the pair of letters G_e and V_e in cipher alphabet (1) of Fig. 6*b*. The letter V_e is the 15th letter to the right of G_e. In cipher alphabet (2), V_e is also the 15th letter to the right of G_e, as is the case in each of the four cipher alphabets in Fig. 6*b*, since the *relative* positions they occupy are the same in each horizontal line in Fig. 6*a*, that is, in each of the successive recordings of the cipher component as the latter is slid to the right against the plain or normal component. If, therefore, the relative positions occupied by two letters, Θ_1 and Θ_2, in such a cipher alphabet, C_1, are known, and if the position of Θ_1 in another cipher alphabet, C_2, belonging to the same series is known, then Θ_2 may at once be placed into its correct position in C_2. Suppose, for example, that as the result of an analysis based upon considerations of frequency, the following values in four cipher alphabets have been tentatively determined:

(1)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					G										Y					V						
(2)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					N										G					P						
(3)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					L										B					I						
(4)	Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher					W										I					Q						

FIGURE 7*a*.

c. The cipher components of these four secondary alphabets may, for convenience, be assembled into a cellular structure, hereinafter called a *sequence reconstruction skeleton*, as shown in Fig. 7*b*. Regarding the top line of the reconstruction skeleton in Fig. 7*b* as being common to all four secondary cipher alphabets listed in Fig. 7*a*, the successive lines of the reconstruction skeleton may now be termed cipher alphabets, and may be referred to by the numbers at the left.

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
Cipher	1					G										Y					V						
	2					N										G					P						
	3					L										B					I						
	4					W										I					Q						

FIGURE 7*b*.

d. The letter G is common to Alphabets 1 and 2. In Alphabet 2 it is noted that N occupies the 10th position to the left of G, and the letter P occupies the 5th position to the right of G. One may therefore place these letters, N and P, in their proper positions in Alphabet 1, the letter N being placed 10 letters before G, and the letter P, 5 letters after G. Thus:

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
	1					G					P					Y					V	N					

Thus, the values of two new letters in Alphabet 1, viz, $P_c=J_p$, and $N_c=U_p$ have been automatically determined; these values were obtained without any analysis based upon the *frequency* of P_c and N_c. Likewise, in Alphabet 2, the letters Y and V may be inserted in these positions:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
2				V	N										G					P					Y	

This gives the new values $V_c=D_p$ and $Y_c=Y_p$ in Alphabet 2. Alphabets 3 and 4 have a common letter I, which permits of the placement of Q and W in Alphabet 3, and of B and L in Alphabet 4.

e. The new values thus found are of course immediately inserted throughout the cryptogram, thus leading to the assumption of further values in the cipher text. This process, *viz, the reconstruction of the primary components*, by the application of the principles of direct symmetry of position to the cells of the reconstruction skeleton, thus facilitates and hastens solution.

f. It must be clearly understood that before the principles of direct symmetry of position can be applied in cases such as the foregoing, *it is necessary that the plain component be a known sequence.* Whether it is the normal sequence or not is immaterial, so long as the sequence is known. Obviously, if the sequence is unknown, symmetry, even if present, cannot be detected by the cryptanalyst because he has no *base* upon which to try out his assumptions for symmetry. In other words, direct symmetry of position is manifested in the illustrative example because the plain component is a known sequence, and not because it is the normal alphabet. The significance of this point will become apparent later on in connection with the problem discussed in Par. 26*b*.

19. Initial steps in the solution of a typical example.—*a.* In the light of the foregoing principles let a typical message now be studied.

MESSAGE

	1	2	3	4	5
A.	Q W B R I	V W Y C A	I S P J L	R B Z E Y	Q W Y E U
B.	L W M G W	I C J C I	M T Z E I	M I B K N	Q W B R I
C.	V W Y I G	B W N B Q	Q C G Q H	I W J K A	G E G X N
D.	I D M R U	V E Z Y G	Q I G V N	C T G Y O	B P D B L
E.	V C G X G	B K Z Z G	I V X C U	N T Z A O	B W F E Q
F.	Q L F C O	M T Y Z T	C C B Y Q	O P D K A	G D G I G
G.	V P W M R	Q I I E W	I C G X G	B L G Q Q	V B G R S
H.	M Y J J Y	Q V F W Y	R W N F L	G X N F W	M C J K X
J.	I D D R U	O P J Q Q	Z R H C N	V W D Y Q	R D G D G
K.	B X D B N	P X F P U	Y X N F G	M P J E L	S A N C D
L.	S E Z Z G	I B E Y U	K D H C A	M B J J F	K I L C J
M.	M F D Z T	C T J R D	M I Y Z Q	A C J R R	S B G Z N
N.	Q Y A H Q	V E D C Q	L X N C L	L V V C S	Q W B I I
P.	I V J R N	W N B R I	V P J E L	T A G D N	I R G Q P
Q.	A T Y E W	C B Y Z T	E V G Q U	V P Y H L	L R Z N Q
R.	X I N B A	I K W J Q	R D Z Y F	K W F Z L	G W F J Q
S.	Q W J Y Q	I B W R X			

b. The principal repetitions of three or more letters have been underlined in the message and the factors (up to 20 only) of the intervals between them are as follows:

QWBRIVWY..... 45=3, 5, 9, 15.
CGXGB..... 60=2, 3, 4, 5, 6, 10, 12, 15, 20.
PJEL..... 95=5, 19.
ZZGI..... 145=5.
BRIV..... 285=3, 5, 15, 19.
BRI..... 45=3, 5, 9, 15.
KAG..... 75=3, 5, 15.
QRD..... 165=3, 5, 15.
QWB..... 45=3, 5, 9, 15.
QWB..... 275=5, 11.
WIC..... 130=2, 5, 10, 13.
XNF..... 45=3, 5, 9, 15.
YZT..... 225=3, 5, 15.
ZTC..... 145=3, 5.

The factor 5 is common to all of these repetitions, and there seems to be every indication that five alphabets are involved. Since the message already appears in groups of five letters, it is unnecessary in this case to rewrite it in groups corresponding to the length of the key. The uniliteral frequency distribution for Alphabet 1 is as follows:

FIGURE 8.

c. Attempts to fit this distribution to the normal on the basis of a direct or reversed standard alphabet do not give positive results, and it is assumed that mixed alphabets are involved. Individual triliteral frequency distributions are then compiled and are shown in Fig. 9. These tables are similar to those made for single mixed alphabet ciphers, and are made in the same way except that instead of taking the letters one after the other, the letters which belong to the separate alphabets now must be assembled in separate tables. For example, in Alphabet 1, the trigraph QAC means that A occurs in Alphabet 1; Q, its prefix, occurs in Alphabet 5, and C, its suffix, occurs in Alphabet 2. All confusion may be avoided by placing numbers indicating the alphabets in which they belong above the letters, thus: $\overset{5\,1\,2}{QAC}$

ALPHABET 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
QC	GW	NT		TV		AE		AS		UD	UW	IT	UT	QP	MX	-W	LB	LA	LA		IW	NW	QI	UX	QR
PT	OP	TC				AD		WC		FI	QX	II		UP		YW	YW	DE			IW				
	GK	TT				LX		HW		FW	LV	OT				NW	QD	RB			UE				
	OW	WB				LW		ND			LR	SY				QC	QD				LC				
	GL							GV				WC				GI					GP				
	GX							WC				GP				QL					QB				
								XD				AB				RI					NW				
								GB				JF				YV					QE				
								IV				DI				NY					IP				
								NR								SW					UP				
								AK								QW									
								QB																	

ALPHABET 2

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
SN	RZ	IJ	IM	GG	MD			MB		IW	QF		WB		BD		ZH	IP	MZ		IX	QB	GN	MJ	
TG	VG	QG	GG	VZ				QG		BZ	BG				OD		IG		CG		QF	VY	BD	QA	
	IE	VG	ID	SZ				QI							VW		LZ		NZ		LV	QY	PF		
	MJ	CB	RG	VD				KL							OJ				MY		IJ	LM	YN		
	SG	IG	KH					MY							MJ				CJ		EG	QB	LN		
	CY	MJ	RZ					XN							VJ				AY			VY			
	IW	AJ													VY							BN			
																						IJ			
																						BF			
																						RN			
																						VD			
																						QB			
																						KF			
																						GF			
																						QJ			

ALPHABET 3

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
YH	WR		PB	BY	WE	CQ	RC	IE	CC		IC	WG	WB		SJ						VC	PM	VC	WC	BE
	IK		PK		LC	EX	DC		WK			DR	WF									KJ		WE	TE
	WR		DR		VW	IV			YJ				XF									BR		WI	EY
	CY		WY		XP	TY			CK				XF											TZ	KZ
	WI		XB		WZ	CX			PQ				AC											IZ	TA
	NR		FZ		WJ	DI			PE				XC											TE	EZ
			EC			CX			BJ				IB											BZ	RN
						LQ			TR															PH	DY
						BR			CR																
						DD			VR																
						BZ			PE																
						AD			WY																
						RQ																			
						VQ																			

ALPHABET 4

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ZO	NQ	YA	GG	ZY	NL	MW	AQ	YG	PL	BN		WR	ZQ		FU	GH	BI				GN	FY	GN	ZG	ZG
	DL	JI	GN	YU	NW		YL	GG	JY	JA						GQ	BI						GG	GO	YT
	DN	XU		ZI	NG			BI	JF	DA						JQ	MU						GG	BQ	ZG
	NA	FO		FQ					WQ	JX						GP	GS							DQ	DT
		HN		IW					FQ							GU	DU							EU	YQ
		ND		JL													JD							ZF	GN
		HA		JL													JR							JQ	YT
		LJ		YW													JN								FL
		DQ															BI								
		NL															WX								
		VS																							

ALPHABET 5

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
CI			CS		JK	IB	QI	RV	CM		JR		KQ	YB	QA	BQ	MQ	RM	ZC	EL		GI	KI	EQ	
KG			RM		YK	YQ		CM			BV		XI	AB		EQ	RS	CQ	ZC	RV		EI	R-	JQ	
KG						XB		EM			FG		VC	CM		YO			ZE	CN		FM		WR	
CM						ZI		RV			ES		CV			QV				RO		EC			
BI						IV		II			CL		BP			QZ				PY					
						XB		RV			ET		ZQ			YR				YK					
						DB					HL		RW			ZA				QV					
						FM					ZG		DI			HV									
						ZI										CL									
																NX									
																JR									
																JQ									
																YI									

Condensed table of repetitions

1-2-3-4-5-1-2-3
Q W B R I V W Y-2

2-3-4-5-1
C G X G B-2

2-3-4-1
P J E L-2

3-4-5-1
B-R-I-V
Z-Z-G-I-2

1-2-3
Q W B-3
V W Y-2

2-3-4
C G X-2
P J E-2
W B R-2
X N F-2

3-4-5
B R I-3
G X G-2
J E L-2
Y Z T-2
Z Z G-2

4-5-1
K A G-2
X G B-2
Z G I-2
Z T C-2
R I V-3

5-1-2
I V W-2
Q R D-2
W I C-2

1-2
Q W-5
V P-3
V W-3

2-3
C G-3
C J-3
P J-3
W B-3
W F-3
W Y-3
X N-3

3-4
B R-3
G Q-4
G X-3
J R-3
N F-3
Y Z-3

4-5
R I-3
Y Q-3
Z T-3

5-1
G B-4
I V-3
Q Q-3

FIGURE 9.

d. One now proceeds to analyze each alphabet distribution, in an endeavor to establish identifications of cipher equivalents. First, of course, attempts should be made to separate the vowels from the consonants in each alphabet, using the same test as in the case of a single mixed-alphabet cipher. There seems to be no doubt about the equivalent of E_p in each alphabet:

$$E = \overset{1}{I}_c, \overset{2}{W}_c, \overset{3}{G}_c, \overset{4}{C}_c, \overset{5}{Q}_c$$

e. The letters of greatest frequency in Alphabet 1 are I, M, Q, V, B, G, L, R, S, and C. $\overset{1}{I}_c$ has already been assumed to be E_p. If $\overset{2}{W}_c$ and $\overset{5}{Q}_c = E_p$, then one should be able to distinguish the vowels from the consonants among the letters M, Q, V, B, G, L, R, S, and C by examining the prefixes of $\overset{2}{W}_c$, and the suffixes of $\overset{5}{Q}_c$. The prefixes and suffixes of these letters, as shown by the triliteral frequency distributions, are these:

Prefixes of $\overset{2}{W}_c$ $(=\overset{2}{E}_p)$	Suffixes of $\overset{5}{Q}_c$ $(=\overset{5}{E}_p)$
Q G K V R B I L	I Q R X L V A Z O

f. Consider now the letter $\overset{1}{M}_c$; it does not occur either as a prefix of $\overset{2}{W}_c$, or as a suffix of $\overset{5}{Q}_c$. Hence it is most probably a vowel, and on account of its high frequency it may be assumed to be O_p. On the other hand, note that $\overset{1}{Q}_c$ occurs five times as a prefix of $\overset{2}{W}_c$ and three times as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, most probably R_p, for it would give the digraph ER $(=\overset{51}{Q}Q_c)$ as occurring three times and RE $(=\overset{12}{Q}W_c)$ as occurring five times.

g. The letter $\overset{1}{V}_c$ occurs three times as a prefix of $\overset{2}{W}_c$ and twice as a suffix of $\overset{5}{Q}_c$. It is therefore a consonant, and on account of its frequency, let it be assumed to be T_p. The letter $\overset{1}{B}_c$ occurs twice as a prefix of $\overset{2}{W}_c$ but not as a suffix of $\overset{5}{Q}_c$. Its frequency is only medium, and it is probably a consonant. In fact, the twice repeated digraph $\overset{12}{B}W_c$ is once a part of the trigraph $\overset{512}{G}BW$, and $\overset{5}{G}_c$, the letter of second highest frequency in Alphabet 5, looks excellent for T_p. Might not the trigraph $\overset{512}{G}BW$ be THE? It will be well to keep this possibility in mind.

h. The letter $\overset{1}{G}_c$ occurs only once as a prefix of $\overset{2}{W}_c$ and does not occur as a suffix of $\overset{5}{Q}_c$. It may be a vowel, but one can not be sure. The letter $\overset{1}{L}_c$ occurs once as a prefix of $\overset{2}{W}_c$ and once as a suffix of $\overset{5}{Q}_c$. It may be considered to be a consonant. $\overset{1}{R}_c$ occurs once as a prefix of $\overset{2}{W}_c$, and twice as a suffix of $\overset{5}{Q}_c$, and is certainly a consonant. Neither the letter $\overset{1}{S}_c$ nor the letter $\overset{1}{C}_c$ occurs as a prefix of $\overset{2}{W}_c$ or as a suffix of $\overset{5}{Q}_c$; both would seem to be vowels, but a study of the prefixes and suffixes of these letters lends more weight to the assumption that $\overset{1}{C}_c$ is a vowel than that $\overset{1}{S}_c$ is a vowel. For all the prefixes of C, viz, $\overset{5}{N}$, $\overset{5}{T}$, and $\overset{5}{W}$, are in subsequent analysis of Alphabet 5 classified as consonants, as are likewise its suffixes, viz, T, C, and B in Alphabet 2. On the other hand, only one prefix, $\overset{5}{L}_c$, and one suffix, $\overset{2}{B}_c$, of $\overset{1}{S}_c$ are later classified as consonants. Since vowels are

more often associated with consonants than with other vowels, it would seem that $\overset{1}{C}_e$ is more likely to be a vowel than $\overset{1}{S}_e$. At any rate $\overset{1}{C}_e$ is assumed to be a vowel, for the present, leaving $\overset{1}{S}_e$ unclassified.

i. Going through the same steps with the remaining alphabets, the following results are obtained:

Alphabet	Consonants	Vowels
1	Q, V, B, L, R, G?	I, M, C.
2	B, C, D, T.	W, P, I.
3	J, N, D, Y, F.	G, Z.
4	Y, Z, J. Q.	C, E?, R?, B?
5	G, N, A, I, W, L, T.	Q, U.

20. Application of principles of direct symmetry of position.—*a*. The next step is to try to determine a few values in each alphabet. In Alphabet 1, from the foregoing analysis, the following data are on hand:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	C?				I				C?						M			Q		V						

Let the values of E_p already assumed in the remaining alphabets, be set down in a reconstruction skeleton, as follows:

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
Cipher	1	C?				I				C?						M			Q		V						
	2					W																					
	3					G																					
	4					C																					
	5					Q																					

FIGURE 10.

b. It is seen that by good fortune the letter Q is common to Alphabets 1 and 5, and the letter C is common to Alphabets 1 and 4. If it is assumed that one is dealing with a case in which a mixed component is sliding against the normal component, one can apply the principles of direct symmetry of position to these alphabets, as outlined in Par. 18. For example, one may insert the following values in Alphabet 5:

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
Cipher	1	C?				I				C?						M			Q		V						
	5		M			Q		V							C?				I				C?				

FIGURE 11.

c. The process at once gives three definite values: $\overset{5}{M}_c=B_p$, $\overset{5}{V}_c=G_p$, $\overset{5}{I}_c=R_p$. Let these deduced values be substantiated by referring to the frequency distribution. Since B and G are normally low or medium frequency letters in plain text, one should find that M_c and V_c, their hypothetical equivalents in Alphabet 5, should have low frequencies. As a matter of fact, they do not appear in this alphabet, which thus far corroborates the assumption. On the other hand, since $\overset{5}{I}_c=R_p$, if the values derived from symmetry of position are correct, $\overset{5}{I}_c$ should be of high frequency, and reference to the distribution shows that I_c is of high frequency. The position of C is doubtful; it belongs either under N_p or V_p. If the former is correct, then the frequency of $\overset{5}{C}_c$ should be high, for it would equal N_p; if the latter is correct, then its frequency should be low, for it would equal V_c. As a matter of fact, $\overset{5}{C}_c$ does not occur, and it must be concluded that it belongs under V_p. This in turn settles the value of $\overset{1}{C}_c$, for it must now be placed definitely under I_p and removed from beneath A_p.

d. The definite placement of C now permits the insertion of new values in Alphabet 4, and one now has the following:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																										
Cipher 1					I				C						M			Q		V						
Cipher 2					W																					
Cipher 3					G																					
Cipher 4	I				C						M			Q		V										
Cipher 5		M			Q		V											I				C				

FIGURE 12.

21. Subsequent steps in solution.—*a.* It is high time that the thus far deduced values, as recorded in the reconstruction skeleton, be inserted in the cipher text, for by this time it must seem that the analysis has certainly gone too far upon unproved hypotheses. The following results are obtained:

MESSAGE

	1					2					3					4					5				
A.	Q	W	B	R	I	V	W	Y	C	A	I	S	P	J	L	R	B	Z	E	Y	Q	W	Y	E	U
	R	E			R	T	E		E		E										R	E			
B.	L	W	M	G	W	I	C	J	C	I	M	T	Z	E	I	M	I	B	K	N	Q	W	B	R	I
		E				E			E	R	O				R	O					R	E			R
C.	V	W	Y	I	G	B	W	N	B	Q	Q	C	G	Q	H	I	W	J	K	A	G	E	G	X	N
	T	E		A			E			E	R		E	N		E	E						E		

D.	I D M R U E	V E Z Y G T	Q I G V N R E P	C T G Y O I E	B P D B L
E.	V C G X G T E	B K Z Z G	I V X C U E E	N T Z A O	B W F E Q E E
F.	Q L F C O R E	M T Y Z T O	C C B Y Q I E	O P D K A	G D G I G E A
G.	V P W M R T K	Q I I E W R	I C G X G E E	B L G Q Q E N E	V B G R S T E
H.	M Y J J Y O	Q V F W Y R	R W N F L E	G X N F W	M C J K X O
J.	I D D R U E	O P J Q Q N E	Z R H C N E	V W D Y Q T E E	R D G D G E
K.	B X D B N	P X F P U	Y X N F G	M P J E L O	S A N C D E
L.	S E Z Z G	I B E Y U E	K D H C A E	M B J J F O	K I L C J E
M.	M F D Z T O	C T J R D I	M I Y Z Q O E	A C J R R	S B G Z N E
N.	Q Y A H Q R E	V E D C Q T E E	L X N C L E	L V V C S E	Q W B I I R E A R
P.	I V J R N E	W N B R I R	V P J E L T	T A G D N E	I R G Q P E E N
Q.	A T Y E W	C B Y Z T I	E V G Q U E N	V P Y H L T	L R Z N Q E
R.	X I N B A	I K W J Q E E	R D Z Y F	K W F Z L E	G W F J Q E E
S.	Q W J Y Q R E E	I B W R X E			

b. The combinations given are excellent throughout and no inconsistencies appear. Note the trigraph $\overset{1\,2\,3}{\text{QWB}}$, which is repeated in the following polygraphs (underlined in the foregoing text):

1	2	3	4	5	1				5	1	2	3	4	5	1
Q	W	B	R	I	V	.	.	.	S	Q	W	B	I	I	I
R	E			R	T	.	.	.		R	E		A	R	E

c. The letter $\overset{3}{B}_c$ is common to both polygraphs, and a little imagination will lead to the assumption of the value $\overset{3}{B}_c = P_p$, yielding the following:

1	2	3	4	5	1				5	1	2	3	4	5	1
Q	W	B	R	I	V	.	.	.	S	Q	W	B	I	I	I
R	E	P	O	R	T	.	.	.	P	R	E	P	A	R	E

d. Note also (in F5) the polygraph $\overset{4}{\underset{A}{I}}\ \overset{5}{G}\ \overset{1}{\underset{T}{V}}\ \overset{2}{P}\ \overset{3}{W}\ \overset{4}{\underset{K}{M}}$, which looks like the word ATTACK. The frequency distributions are consulted to see whether the frequencies given for $\overset{5}{G}_c$ and $\overset{2}{P}_c$ are high enough for T_p and A_p, respectively, and also whether the frequency of $\overset{3}{W}_c$ is good enough for C_p; it is noted that they are excellent. Moreover, the digraph $\overset{5\,1}{GB}_c$, which occurs four times, looks like TH, thus making $\overset{1}{B}_c = H_p$. Does the insertion of these four new values in our diagram of alphabets bring forth any inconsistencies? The insertion of the value $\overset{2}{P}_c = A_p$ and $\overset{1}{B}_c = H_p$ gives no indications either way, since neither letter has yet been located in any of the other alphabets. The insertion of the value $\overset{5}{G}_c = T_p$ gives a value common to Alphabets 3 and 5, for the value $\overset{3}{G}_c = E_p$ was assumed long ago. Unfortunately an inconsistency is found here. The letter I has been placed two letters to the left of G in the mixed component, and has given good results in Alphabets 1 and 5; if the value $\overset{3}{W}_c = C_p$ (obtained above from the assumption of the word ATTACK) is correct, then W, and not I, should be the second letter to the left of G. Which shall be retained? There has been so far nothing to establish the value of $\overset{3}{G}_c = E_p$; this value was assumed from frequency considerations solely. Perhaps it is wrong. It certainly behaves like a vowel, and one may see what happens when one changes its value to O_p. The following placements in the reconstruction skeleton result from the analysis, when only two or three new values have been added as a result of the clues afforded by the deductions:

		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain																											
Cipher	1			S		I		G	B	C						M		P	Q	R	V	W					
	2	P	Q	R	V	W								S		I		G	B	C						M	
	3	R	V	W								S		I		G	B	C						M		P	Q
	4	I		G	B	C						M		P	Q	R	V	W								S	
	5		M		P	Q	R	V	W								S		I		G	B	C				

FIGURE 13*a*.

e. Many new values are produced, and these are inserted throughout the message, yielding the following:

	1	2	3	4	5
A.	Q W B R I R E P O R	V W Y C A T E E	I S P J L E M Y	R B Z E Y S R	Q W Y E U R E
B.	L W M G W E W C H	I C J C I E S E R	M T Z E I O R	M I B K N O O P	Q W B R I R E P O R
C.	V W Y I G T E A T	B W N B Q H E D E	Q C G Q H R S O N	I W J K A E E	G E G X N G O
D.	I D M R U E W O	V E Z Y G T T	Q I G V N R O O P	C T G Y O I O	B P D B L H A D
E.	V C G X G T S O T	B K Z Z G H T	I V X C U E D E	N T Z A O	B W F E Q H E E
F.	Q L F C O R E	M T Y Z T O	C C B Y Q I S P E	O P D K A A	G D G I G G O A T
G.	V P W M R T A C K F	Q I I E W R O M H	I C G X G E S O T	B L G Q Q H O N E	V B G R S T R O O P
H.	M Y J J Y O	Q V F W Y R D Q	R W N F L S E	G X N F W G H	M C J K X O S
J.	I D D R U E O	O P J Q Q A N E	Z R H C N C E	V W D Y Q T E E	R D G D G S O T
K.	B X D B N H D	P X F P U Q M	Y X N F G T	M P J E L O A	S A N C D C E
L.	S E Z Z G C T	I B E Y U E R	K D H C A E	M B J J F O R	K I L C J O E
M.	M F D Z T O	C T J R D I O	M I Y Z Q O O E	A C J R R S O F	S B G Z N C R O
N.	Q Y A H Q R E	V E D C Q T . E E	L X N C L E	L V V C S D B E P	Q W B I I R E P A R
P.	I V J R N E D O	W N B R I U P O R	V P J E L T A	T A G D N O	I R G Q P E C O N D
Q.	A T Y E W H	C B Y Z T I R	E V G Q U D O N	V P Y H L T A	L R Z N Q C E
R.	X I N B A O D	I K W J Q E E	R D Z Y F S	K W F Z L E	G W F J Q G E E
S.	Q W J Y Q R E E	I B W R X E R O			

22. Completing the solution.—*a.* Completion of solution is now a very easy matter. The mixed component is finally found to be the following sequence, based upon the word EXHAUSTING:

E X H A U S T I N G B C D F J K L M O P Q R V W Y Z

and the completely reconstructed skeleton of the cipher square is shown in Fig. 13*b*.

Plain		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	1	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H
	2	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O
	3	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q
	4	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T
	5	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K

FIGURE 13*b*.

b. Note that the successive equivalents of A_p spell the word APRIL, which is the key for the message. The plain-text message is as follows:

REPORTED ENEMY HAS RETIRED TO NEWCHESTER. ONE TROOP IS REPORTED AT HENDERSON MEETING HOUSE: TWO OTHER TROOPS IN ORCHARD AT SOUTHWEST EDGE OF NEWCHESTER. 2D SQ IS PREPARING TO ATTACK FROM THE SOUTH. ONE TROOP OF 3D SQ IS ENGAGING HOSTILE TROOP AT NEWCHESTER. REST OF 3D SQ IS MOVING TO ATTACK NEWCHESTER FROM THE NORTH. MOVE YOUR SQ INTO WOODS EAST OF CROSSROAD 539 AND BE PREPARED TO SUPPORT ATTACK OF 2D AND 3D SQ. DO NOT ADVANCE BEYOND NEWCHESTER. MESSAGES HERE.

TREER,
COL.

c. The preceding case is a good example of the value of the principles of direct symmetry of position when applied properly to a cryptogram enciphered by the sliding of a mixed component against the normal. The cryptanalyst starts off with only a very limited number of assumptions and builds up many new values as a result of the placement of the few original values in the reconstruction skeleton.

23. Solution of subsequent messages enciphered by the same cipher component.—*a. Preliminary remarks.*—Let it be supposed that the correspondents are using the same basic or primary component but with different key words for other messages. Can the knowledge of the sequence of letters in the reconstructed primary component be used to solve the subsequent messages? It has been shown that in the case of a monoalphabetic cipher in which a mixed alphabet was used, the process of completing the plain component could be applied to solve subsequent messages in which the same cipher component was used, even though the cipher component was set at a different key letter. A modification of the procedure used in that case can be used in this case, where a plurality of cipher alphabets based upon a sliding primary component is used.

b. The message.—Let it be supposed that the following message passing between the same two correspondents as in the preceding message has been intercepted:

MESSAGE

SFDZR YRRKX MIWLL AQRLU RQFRT IJQKF XUWBS MDJZK
MICQC UDPTV TYRNH TRORV BQLTI QBNPR RTUHD PTIVE
RMGQN LRATQ PLUKR KGRZF JCMGP IHSMR GQRFX BCABA
OEMTL PCXJM RGQSZ VB

c. Factoring and conversion into plain component equivalents.—The presence of a repetition of a four-letter polygraph whose interval is 21 letters suggests a key word of seven letters. There are very few other repetitions, and this is to be expected in a short message with a key of such length.

1	2	3	4	5	6	7
S	F	D	Z	R	Y	R
R	K	X	M	I	W	L
L	A	Q	R	L	U	R
Q	F	R	T	I	J	Q
K	F	X	U	W	B	S
M	D	J	Z	K	M	I
C	Q	C	U	D	P	T
V	T	Y	R	N	H	T
R	O	R	V	B	Q	L
T	I	Q	B	N	P	R
R	T	U	H	D	P	T
I	V	E	R	M	G	Q
N	L	R	A	T	Q	P
L	U	K	R	K	G	R
Z	F	J	C	M	G	P
I	H	S	M	R	G	Q
R	F	X	B	C	A	B
A	O	E	M	T	L	P
C	X	J	M	R	G	Q
S	Z	V	B			

FIGURE 14.

d. Transcription into periods.—Let the message be written in groups of seven letters, in columnar fashion, as shown in Fig. 14. The letters in each column belong to a single alphabet. Let the letters in each column be converted into their plain-component equivalents by setting the reconstructed cipher component against the normal alphabet at any arbitrarily selected point, for example, that shown below:

1	2	3	4	5	6	7
F	N	M	Z	V	Y	V
V	P	B	R	H	X	Q
Q	D	U	V	Q	E	V
U	N	V	G	H	O	U
P	N	B	E	X	K	F
R	M	O	Z	P	R	H
L	U	L	E	M	T	G
W	G	Y	V	I	C	G
V	S	V	W	K	U	Q
G	H	U	K	I	T	V
V	G	E	C	M	T	G
H	W	A	V	R	J	U
I	Q	V	D	G	U	T
Q	E	P	V	P	J	V
Z	N	O	L	R	J	T
H	C	F	R	V	J	U
V	N	B	K	L	D	K
D	S	A	R	G	Q	T
L	B	O	R	V	J	U
F	Z	W	K			

FIGURE 15.

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z

The columns of equivalents are now as shown in Fig. 15.

e. Examination and selection of generatrices.—It has been shown that in the case of a monoalphabetic cipher it was merely necessary to complete the normal alphabet sequence beneath the plain-component equivalents and the plain text all reappeared on one generatrix. It was also found that in the case of a multiple-alphabet cipher involving standard alphabets, the plain-text equivalents of each alphabet reappeared on the same generatrix, and it was necessary only to combine the proper generatrices in order to produce the plain text of the message. In the case at hand both processes are combined: the normal alphabet sequence is continued beneath the letters of each column and then the generatrices are combined to produce the plain text. The completely developed generatrix diagrams for the first two columns are as follows (Fig. 16):

	Column 1	Column 2
	FVQUPRLWVGVHIQZHVDLF	NPDNNMUGSHGWQENCNSBZ
1	GWRVQSMXWHWIJRAIWEMG	OQEOONVHTIHXRFODOTCA
2	HXSWRTNYXIXJKSBJXFNH	PRFPPOWIUJIYSGPEPUDB
3	IYTXSUOZYJYKLTCKYGOI	QSGQQPXJVKJZTHQFQVEC
4	JZUYTVPAZKZLMUDLZHPJ	RTHRRQYKWLKAUIRGRWFD
5	KAVZUWQBALAMNVEMAIQK	SUISSRZLXMLBVJSHSXGE
6	LBWAVXRCBMBNOWFNBJRL	TVJTTSAMYNMCWKTITYHF
7	MCXBWYSDCNCOPXGOCKSM	UWKUUTBNZONDXLUJUZIG
8	NDYCXZTEDODPQYHPDLTN	VXLVVUCOAPOEYMVKVAJH
9	OEZDYAUFEPEQRZIQEMUO	WYMWWVDPBQPFZNWLWBKI
10	PFAEZBVGFQFRSAJRFNVP	XZNXXWEQCRQGAOXMXCLJ
11	QGBFACWHGRGSTBKSGOWQ	YAOYYXFRDSRHBPYNYDMK
12	RHCGBDXIHSHTUCLTHPXR	ZBPZZYGSETSICQZOZENL
13	SIDHCEYJITIUVDMUIQYS	ACQAAZHTFUTJDRAPAFOM
14	TJEIDFZKJUJVWENVJRZT	BDRBBAIUGVUKESBQBGPN
15	UKFJEGALKVKWXFOWKSAU	CESCCBJVHWVLFTCRCHQO
16	VLGKFHBMLWLXYGPXLTBV	DFTDDCKWIXWMGUDSDIRP
17	WMHLGICNMXMYZHQYMUCW	EGUEEDLXJYXNHVETEJSQ
18	XNIMHJDONYNZAIRZNVDX	FHVFFEMYKZYOIWFUFKTR
19	YOJNIKEPOZOABJSAOWEY	GIWGGFNZLAZPJXGVGLUS
20	ZPKOJLFQPAPBCKTBPXFZ	HJXHHGOAMBAQKYHWHMVT
21	AQLPKMGRQBQCDLUCQYGA	IKYIIHPBNCBRLZIXINWU
22	BRMQLNHSRCRDEMVDRZHB	JLZJJIQCODCSMAJYJOXV
23	CSNRMOITSDSEFNWESAIC	KMAKKJRDPEDTNBKZKPYW
24	DTOSNPJUTETFGOXFTBJD	LNBLLKSEQFEUOCLALQZX
25	EUPTOQKVUFUGHPYGUCKE	MOCMMLTFRGFVPDMBMRAY

Figure 16.

1	2
C	O
S	Q
N	E
R	O
M	O
O	N
I	V
T	H
S	T
D	I
S	H
E	X
F	R
N	F
W	O
E	D
S	O
A	T
I	C
C	A

Figure 17a.

f. Combining the selected generatrices.—After some experimenting with these generatrices the 23d generatrix of Column 1 and the 1st of Column 2, which yield the digraphs shown in Fig. 17*a*, are combined. The generatrices of the subsequent columns are examined to select those which may be added to these already selected in order to build up the plain text. The results are shown in Fig. 17*b*. This process is a very valuable aid in the solution of messages after the primary component has been recovered as a result of the longer and more detailed analysis of the frequency distributions of the first message intercepted. Very often a short message can be solved in no other way than the one shown, if the primary component is completely known.

g. Recovery of the key.—It may be of interest to find the key word for the message. Assuming that enciphering method number 1 (see Par. 7*f*, page 6) were known to be employed, all that is necessary is to set the mixed component of the cipher alphabet underneath the plain component so as to produce the cipher letter indicated as the equivalent of any given plain-text letter in each of the alphabets. For example, in the first alphabet it is noted that $C_p = S_c$. Adjust the two components under each other so as to bring S of the cipher component beneath C of the plain component, thus:

1	2	3	4	5	6	7
C	O	F	I	R	S	T
S	Q	U	A	D	R	O
N	E	N	E	M	Y	T
R	O	O	P	D	I	S
M	O	U	N	T	E	D
O	N	H	I	L	L	F
I	V	E	N	I	N	E
T	H	R	E	E	W	E
S	T	O	F	G	O	O
D	I	N	T	E	N	T
S	H	X	L	I	N	E
E	X	T	E	N	D	S
F	R	O	M	C	O	R
N	F	I	E	L	D	T
W	O	H	U	N	D	R
E	D	Y	A	R	D	S
S	O	U	T	H	X	I
A	T	T	A	C	K	R
I	C	H	A	R	D	S
C	A	P	T			

Figure 17b.

Plain..... ABCDEFGHIJKLMNOPQRSTUVWXYZABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher..... EXHAUSTINGBCDFJKLMOPQRVWYZ

It is noted that $A_p = A_c$. Hence, the first letter of the key word to the message is A. The 2d, 3d, 4th, . . . 7th key letters are found in exactly the same manner, and the following is obtained:

When C O F I R S T equals
S F D Z R Y R then A_p successively equals
A Z I M U T H

24. Summation of relative frequencies as an aid to the selection of the correct generatrices.—*a.* In the foregoing example, under subparagraph *f*, there occurs this phrase: "After some experimenting with these generatrices . . ." By this was meant, of course, that the selection of the correct initial pair of generatrices of plain-text equivalents is in this process a matter of trial and error. The test of "correctness" is whether, when juxtaposed, the two generatrices so selected yield "good" digraphs, that is, high-frequency digraphs such as occur in normal plain text. In his early efforts the student may have some difficulty in selecting, merely by ocular examination, the most likely generatrices to try. There may be in each diagram several generatrices which contain good assortments of high-frequency letters, and the number of trials of combinations of generatrices may be quite large. Perhaps a simple mathematical method may be of assistance in the process.

b. Suppose, in Fig. 16, that each letter were accompanied by a number which corresponds to its relative frequency in normal English telegraphic text. Then, by adding the numbers along each *horizontal* line, the totals thus obtained will serve as relative numerical measures of the frequency values of the respective generatrices. Theoretically, the generatrix with the greatest value will be the correct generatrix because its total will represent the sum of the individual values of the actual plaintext letters. In actual practice, of course, the generatrix with the greatest value may not be the correct one, but the correct one will certainly be among the three or four generatrices with the largest values. Thus, the number of trials may be greatly reduced, in the attempt to put together the correct generatrices.

c. Using the preceding message as an example, note the respective generatrix values in Fig. 18. The frequency values of the respective letters shown in the figure are based upon the normal distribution for War Department telegraphic text (see Table 3, Appendix 1, Military Cryptanalysis, Part I).

Column 1

Generatrix																				Frequency value
0	F	V	Q	U	P	R	L	W	V	G	V	H	I	Q	Z	H	V	D	L	F	57
	3	2	0	3	3	8	4	2	2	2	2	3	7	0	0	3	2	4	4	3	
1	G	W	R	V	Q	S	M	X	W	H	W	I	J	R	A	I	W	E	M	G	77
	2	2	8	2	0	6	2	0	2	8	2	7	0	8	7	7	2	13	2	2	
2	H	X	S	W	R	T	N	Y	X	I	X	J	K	S	B	J	X	F	N	H	66
	3	0	6	2	8	9	8	2	0	7	0	0	0	6	1	0	0	3	8	3	
3	I	Y	T	X	S	U	O	Z	Y	J	Y	K	L	T	C	K	Y	G	O	I	74
	7	2	9	0	6	3	8	0	2	0	2	0	4	9	3	0	2	2	8	7	
4	J	Z	U	Y	T	V	P	A	Z	K	Z	L	M	U	D	L	Z	H	P	J	49
	0	0	3	2	9	2	3	7	0	0	0	4	2	3	4	4	0	3	3	0	
5	K	A	V	Z	U	W	Q	B	A	L	A	M	N	V	E	M	A	I	Q	K	74
	0	7	2	0	3	2	0	1	7	4	7	2	8	2	13	2	7	7	0	0	
6	L	B	W	A	V	X	R	C	B	M	B	N	O	W	F	N	B	J	R	L	73
	4	1	2	7	2	0	8	3	1	2	1	8	8	2	3	8	1	0	8	4	
7	M	C	X	B	W	Y	S	D	C	N	C	O	P	X	G	O	C	K	S	M	66
	2	3	0	1	2	2	6	4	3	8	3	8	3	0	2	8	3	0	6	2	
8	N	D	Y	C	X	Z	T	E	D	O	D	P	Q	Y	H	P	D	L	T	N	91
	8	4	2	3	0	0	9	13	4	8	4	3	0	2	3	3	4	4	9	8	
9	O	E	Z	D	Y	A	U	F	E	P	E	Q	R	Z	I	Q	E	M	U	O	110
	8	13	0	4	2	7	3	3	13	3	13	0	8	0	7	0	13	2	3	8	
10	P	F	A	E	Z	B	V	G	F	Q	F	R	S	A	J	R	F	N	V	P	82
	3	3	7	13	0	1	2	2	3	0	3	8	6	7	0	8	3	8	2	3	
11	Q	G	B	F	A	C	W	H	G	R	G	S	T	B	K	S	G	O	W	Q	67
	0	2	1	3	7	3	2	3	2	8	2	6	9	1	0	6	2	8	2	0	
12	R	H	C	G	B	D	X	I	H	S	H	T	U	C	L	T	H	P	X	R	82
	8	3	3	2	1	4	0	7	3	6	3	9	3	3	4	9	3	3	0	8	
13	S	I	D	H	C	E	Y	J	I	T	I	U	V	D	M	U	I	Q	Y	S	90
	6	7	4	3	3	13	2	0	7	9	7	3	2	4	2	3	7	0	2	6	
14	T	J	E	I	D	F	Z	K	J	U	J	V	W	E	N	V	J	R	Z	T	83
	9	0	13	7	4	3	0	0	0	3	0	2	2	13	8	2	0	8	0	9	
15	U	K	F	J	E	G	A	L	K	V	K	W	X	F	O	W	K	S	A	U	65
	3	0	3	0	13	2	7	4	0	2	0	2	0	3	8	2	0	6	7	3	
16	V	L	G	K	F	H	B	M	L	W	L	X	Y	G	P	X	L	T	B	V	50
	2	4	2	0	3	3	1	2	4	2	4	0	2	2	3	0	4	9	1	2	
17	W	M	H	L	G	I	C	N	M	X	M	Y	Z	H	Q	Y	M	U	C	W	52
	2	2	3	4	2	7	3	8	2	0	2	2	0	3	0	2	2	3	3	2	
18	X	N	I	M	H	J	D	O	N	Y	N	Z	A	I	R	Z	N	V	D	X	86
	0	8	7	2	3	0	4	8	8	2	8	0	7	7	8	0	8	2	4	0	
19	Y	O	J	N	I	K	E	P	O	Z	O	A	B	J	S	A	O	W	E	Y	103
	2	8	0	8	7	0	13	3	8	0	8	7	1	0	6	7	8	2	13	2	
20	Z	P	K	O	J	L	F	Q	P	A	P	B	C	K	T	B	P	X	F	Z	51
	0	3	0	8	0	4	3	0	3	7	3	1	3	0	9	1	3	0	3	0	
21	A	Q	L	P	K	M	G	R	Q	B	Q	C	D	L	U	C	Q	Y	G	A	55
	7	0	4	3	0	2	2	8	0	1	0	3	4	4	3	3	0	2	2	7	
22	B	R	M	Q	L	N	H	S	R	C	R	D	E	M	V	D	R	Z	H	B	88
	1	8	2	0	4	8	3	6	8	3	8	4	13	2	2	4	8	0	3	1	
23	C	S	N	R	M	O	I	T	S	D	S	E	F	N	W	E	S	A	I	C	129
	3	6	8	8	2	8	7	9	6	4	6	13	3	8	2	13	6	7	7	3	
24	D	T	O	S	N	P	J	U	T	E	T	F	G	O	X	F	T	B	J	D	102
	4	9	8	6	8	3	0	3	9	13	9	3	2	8	0	3	9	1	0	4	
25	E	U	P	T	O	Q	K	V	U	F	U	G	H	P	Y	G	U	C	K	E	78
	13	3	3	9	8	0	0	2	3	3	3	2	3	3	2	2	3	3	0	13	

COLUMN 2

Generatrix																				Frequency value
0	N	P	D	N	N	M	U	G	S	H	G	W	Q	E	N	C	N	S	B	Z	90
	8	3	4	8	8	2	3	2	6	3	2	2	0	13	8	3	8	6	1	0	
1	O	Q	E	O	O	N	V	H	T	I	H	X	R	F	O	D	O	T	C	A	119
	8	0	13	8	8	8	2	3	9	7	3	0	8	3	8	4	8	9	3	7	
2	P	R	F	P	P	O	W	I	U	J	I	Y	S	G	P	E	P	U	D	B	84
	3	8	3	3	3	8	2	7	3	0	7	2	6	2	3	13	3	3	4	1	
3	Q	S	G	Q	Q	P	X	J	V	K	J	Z	T	H	Q	F	Q	V	E	C	46
	0	6	2	0	0	3	0	0	2	0	0	0	9	3	0	3	0	2	13	3	
4	R	T	H	R	R	Q	Y	K	W	L	K	A	U	I	R	G	R	W	F	D	88
	8	9	3	8	8	0	2	0	2	4	0	7	3	7	8	2	8	2	3	4	
5	S	U	I	S	S	R	Z	L	X	M	L	B	V	J	S	H	S	X	G	E	79
	6	3	7	6	6	8	0	4	0	2	4	1	2	0	6	3	6	0	2	13	
6	T	V	J	T	T	S	A	M	Y	N	M	C	W	K	T	I	T	Y	H	F	94
	9	2	0	9	9	6	7	2	2	8	2	3	2	0	9	7	9	2	3	3	
7	U	W	K	U	U	T	B	N	Z	O	N	D	X	L	U	J	U	Z	I	G	68
	3	2	0	3	3	9	1	8	0	8	8	4	0	4	3	0	3	0	7	2	
8	V	X	L	V	V	U	C	O	A	P	O	E	Y	M	V	K	V	A	J	H	73
	2	0	4	2	2	3	3	8	7	3	8	13	2	2	2	0	2	7	0	3	
9	W	Y	M	W	W	V	D	P	B	Q	P	F	Z	N	W	L	W	B	K	I	50
	2	2	2	2	2	2	4	3	1	0	3	3	0	8	2	4	2	1	0	7	
10	X	Z	N	X	X	W	E	Q	C	R	Q	G	A	O	X	M	X	C	L	J	60
	0	0	8	0	0	2	13	0	3	8	0	2	7	8	0	2	0	3	4	0	
11	Y	A	O	Y	Y	X	F	R	D	S	R	H	B	P	Y	N	Y	D	M	K	75
	2	7	8	2	2	0	3	8	4	6	8	3	1	3	2	8	2	4	2	0	
12	Z	B	P	Z	Z	Y	G	S	E	T	S	I	C	Q	Z	O	Z	E	N	L	85
	0	1	3	0	0	2	2	6	13	9	6	7	3	0	0	8	0	13	8	4	
13	A	C	Q	A	A	Z	H	T	F	U	T	J	D	R	A	P	A	F	O	M	93
	7	3	0	7	7	0	8	9	3	3	9	0	4	8	7	3	7	3	8	2	
14	B	D	R	B	B	A	I	U	G	V	U	K	E	S	B	Q	B	G	P	N	73
	1	4	8	1	1	7	7	3	2	2	3	0	13	6	1	0	1	2	3	8	
15	C	E	S	C	C	B	J	V	H	W	V	L	F	T	C	R	C	H	Q	O	79
	3	13	6	3	3	1	0	2	3	2	2	4	3	9	3	8	3	3	0	8	
16	D	F	T	D	D	C	K	W	I	X	W	M	G	U	D	S	D	I	R	P	77
	4	3	9	4	4	3	0	2	7	0	2	2	2	3	4	6	4	7	8	3	
17	E	G	U	E	E	D	L	X	J	Y	X	N	H	V	E	T	E	J	S	Q	108
	13	2	3	13	13	4	4	0	0	2	0	8	3	2	13	9	13	0	6	0	
18	F	H	V	F	F	E	M	Y	K	Z	Y	O	I	W	F	U	F	K	T	R	76
	3	3	2	3	3	13	2	2	0	0	2	8	7	2	3	3	3	0	9	8	
19	G	I	W	G	G	F	N	Z	L	A	Z	P	J	X	G	V	G	L	U	S	59
	2	7	2	2	2	3	8	0	4	7	0	3	0	0	2	2	2	4	3	6	
20	H	J	X	H	H	G	O	A	M	B	A	Q	K	Y	H	W	H	M	V	T	59
	3	0	0	3	3	2	8	7	2	1	7	0	0	2	3	2	3	2	2	9	
21	I	K	Y	I	I	H	P	B	N	C	B	R	L	Z	I	X	I	N	W	U	81
	7	0	2	7	7	3	3	1	8	3	1	8	4	0	7	0	7	8	2	3	
22	J	L	Z	J	J	I	Q	C	O	D	C	S	M	A	J	Y	J	O	X	V	56
	0	4	0	0	0	7	0	3	8	4	3	6	2	7	0	2	0	8	0	2	
23	K	M	A	K	K	J	R	D	P	E	D	T	N	B	K	Z	K	P	Y	W	66
	0	2	7	0	0	0	8	4	3	13	4	9	8	1	0	0	0	3	2	2	
24	L	N	B	L	L	K	S	E	Q	F	E	U	O	C	L	A	L	Q	Z	X	85
	4	8	1	4	4	0	6	13	0	3	13	3	8	3	4	7	4	0	0	0	
25	M	O	C	M	M	L	T	F	R	G	F	V	P	D	M	B	M	R	A	Y	77
	2	8	3	2	2	4	9	3	8	2	3	2	3	4	2	1	2	8	7	2	

FIGURE 18.

d. It will be noted that the frequency value of the 23d generatrix for the first column of cipher letters is the greatest; that of the first generatrix for the second column is the greatest. In both cases these are the correct generatrices. Thus the selection of the correct generatrices in such cases has been reduced to a purely mathematical basis which is at times of much assistance in effecting a quick solution. Moreover, an understanding of the principles involved will be of considerable value in subsequent work.

25. Solution by the probable-word method.—*a.* Occasionally one may encounter a cryptogram which is so short that it contains no recurrences even of digraphs, and thus gives no indications of the number of alphabets involved. If the sliding mixed component is known, one may apply the method illustrated in Par. 15, assuming the presence of a probable word, checking it against the text and the sliding components to establish a key, if the correspondents are using key words.

b. For example, suppose that the presence of the word ENEMY is assumed in the message in Par. 23*b* above. One proceeds to check it against an unknown key word, sliding the already reconstructed mixed component against the normal and starting with the first letter of the cryptogram, in this manner:

When ENEMY equals
SFDZR then A_p successively equals
XENFW

The sequence XENFW spells no intelligible word. Therefore, the location of the assumed word ENEMY is shifted one letter forward in the cipher text, and the test is made again, just as was explained in Par. 15. When the group AQRLU is tried, the key letters ZIMUT are obtained, which, taken as a part of a word, suggests the word AZIMUTH. The method must yield solution when the correct assumptions are made.

c. The danger to cryptographic security resulting from the inclusion of *cryptographed* addresses and signatures in cryptographic messages becomes quite obvious in the light of solution by the probable-word method. To illustrate, reference is made to the message employed in Pars. 19–22. It will be noted in Par. 22*b* that the message carried a signature (Treer, Col.) and that the latter was enciphered. Suppose that this were an authorized practice, and that every message could be assumed to conclude with a cryptographed signature. The signature "TREER COL" would at once afford a very good basis for the quick solution of subsequent messages emanating from the same headquarters as did the first message, because presumably this same signature would appear in other messages. It is for this reason that addresses and signatures must not be cryptographed; if they must be included they should be cryptographed in a totally different system or by a wholly different method, perhaps by means of a special address and signature code. It would be best, however, to omit all addresses and signatures, and to let the call signs of the headquarters concerned also convey these parts of the message, leaving the delivery to the addressee a matter for local action.

26. Solution when the plain component is a mixed sequence, the cipher component, the normal.—*a.* This falls under Case B (2) outlined in Par. 6. It is not the usual method of employing a single mixed component, but may be encountered occasionally in cipher devices.

b. The preliminary steps, as regards factoring to determine the length of the period, are the same as usual. The message is then transcribed into its periods. Frequency distributions are then made, as usual, and these are attacked by the principles of frequency and recurrence. An attempt is made to apply the principles of direct symmetry of position, but this attempt will be futile, for the reason that the plain component is in this case an *unknown* mixed sequence.

(See Par. 18*d*.) Any attempt to find symmetry in the secondary alphabets based upon the normal sequence can therefore disclose no symmetry because the symmetry which exists is based upon a wholly different sequence.

c. However, if the principles of direct symmetry of position are of no avail in this case, there are certain other principles of symmetry which may be employed to great advantage. To explain them an actual example will be used. Let it be assumed that it is known to the cryptanalyst that the enemy is using the general system under discussion, *viz*, a mixed sequence, variable from day to day, is used as plain component; the normal sequence is used as cipher component; and a repeating key, variable from message to message, is used in the ordinary manner.

The following message has been intercepted:

	1	2	3	4	5	6
A.	Q E O V K	L R M L Z	J V G T G	N D L V K	E V N T Y	E R M U E
B.	V R Z M O	Y A A M P	D K E I J	S F M Y O	Y H M M E	G Q A M B
C.	U Q A X R	H U F B U	K Q Y M U	N E L V T	K Q I L E	K Z B U E
D.	U L I B K	N D A X B	X U D G L	L A D V K	P O A Y O	D K K Y K
E.	L A D H Y	B V N F V	U E E M E	F F M T E	G V W B Y	T V D Z L
F.	S P B H B	X V A Z C	U D Y U E	L K M M A	E U D D K	N C F S H
G.	H S A H Y	T M G U J	H Q X P P	D K O U E	X U Q V B	F V W B X
H.	N X A L B	T C D L M	I V A A A	N S Z I L	O V W V P	Y A G Z L
J.	S H M M E	G Q D H O	Y H I V P	N C R R E	X K D Q Z	G K N C G
K.	N Q G U Y	J I W Y Y	T M A H W	X R L B L	O A D L G	N Q G U Y
L.	J U U G B	J H R V X	E R F L E	G W G U O	X E D T P	D K E I Z
M.	V X N W A	F A A N E	M K G H B	S S N L O	K J C B Z	T G G L O
N.	P K M B X	H G E R Y	T M W L Z	N Q C Y Y	T M W I P	D K A T E
P.	F L N U J	N D T V X	J R Z T L	O P A H C	D F Z Y Y	D E Y C L
Q.	G P G T Y	T E C X B	H Q E B R	K V W M U	N I N G J	I Q D L P
R.	J K A T E	G U W B R	H U Q W M	V R Q B W	Y R F B F	K M W M B
S.	T M U L Z	L A A H Y	J G D V K	L K R R E	X K N A O	N D S B X
T.	X C G Z A	H D G T L	V K M B W	I S A U E	F D N W P	N L Z I J
V.	S R Q Z L	A V N H L	G V W V K	F I G H P	G E C Z U	K Q A P

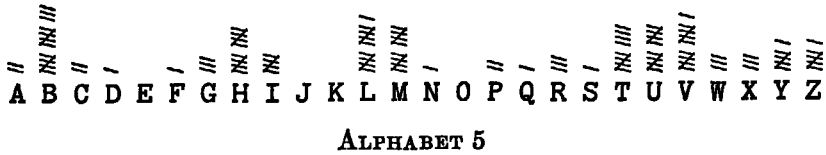
d. A study of the recurrences and factoring their intervals discloses that five alphabets are involved. Uniliteral frequency distributions are made and are shown in Fig. 19*a*:

ALPHABET 1

ALPHABET 2

ALPHABET 3

ALPHABET 4



ALPHABET 5

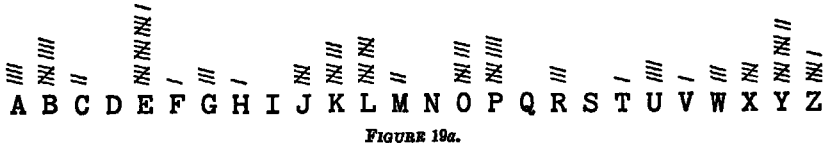


FIGURE 19a.

e. Since the cipher component in this case is the normal alphabet, *it follows that the five frequency distributions are based upon a sequence which is known, and therefore, the five frequency distributions should manifest a direct symmetry of distribution of crests and troughs.* By virtue of this symmetry and by shifting the five distributions relative to one another to proper superimpositions, the several distributions may be combined into a single uniliteral distribution. Note how this shifting has been done in the case of the five illustrative distributions:

ALPHABET 1

ALPHABET 2

ALPHABET 3

ALPHABET 4

ALPHABET 5

FIGURE 19b.

f. The superimposition of the respective distributions enables one to convert the cipher letters of the five alphabets into one alphabet. Suppose it is decided to convert Alphabets 2, 3, 4, and 5 into Alphabet 1. It is merely necessary to substitute for the respective letters in the four alphabets those which stand above them in Alphabet 1. For example, in Fig. 19*b*, X_c in Alphabet 2 is directly under A_c in Alphabet 1; hence, if the superimposition is correct then $\overset{2}{X_c}=\overset{1}{A_c}$. Therefore, in the cryptogram it is merely necessary to replace every X_c in the second position by A_c. Again T_c in Alphabet 3=A_c in Alphabet 1; therefore, in the cryptogram one replaces every T_c in the third position by A_c. The entire process, hereinafter designated as *conversion into monoalphabetic terms*, gives the following *converted* message:

	1					2					3					4					5					6				
A.	Q	H	V	H	T	L	U	T	X	I	J	Y	N	F	P	N	G	S	H	T	E	Y	U	F	H	E	U	T	G	N
B.	V	U	G	Y	X	Y	D	H	Y	Y	D	N	L	U	S	S	I	T	K	X	Y	K	T	Y	N	G	T	H	Y	K
C.	U	T	H	J	A	H	X	M	N	D	K	T	F	Y	D	N	H	S	H	C	K	T	P	X	N	K	C	I	G	N
D.	U	O	P	N	T	N	G	H	J	K	X	X	K	S	U	L	D	K	H	T	P	R	H	K	X	D	N	R	K	T
E.	L	D	K	T	H	B	Y	U	R	E	U	H	L	Y	N	F	I	T	F	N	G	Y	D	N	H	T	Y	K	L	U
F.	S	S	I	T	K	X	Y	H	L	L	U	G	F	G	N	L	N	T	Y	J	E	X	K	P	T	N	F	M	E	Q
G.	H	V	H	T	H	T	P	N	G	S	H	T	E	B	Y	D	N	V	G	N	X	X	X	H	K	F	Y	D	N	G
H.	N	A	H	X	K	T	F	K	X	V	I	Y	H	M	J	N	V	G	U	U	O	Y	D	H	Y	Y	D	N	L	U
J.	S	K	T	Y	N	G	T	K	T	X	Y	K	P	H	Y	N	F	Y	D	N	X	N	K	C	I	G	N	U	O	P
K.	N	T	N	G	H	J	L	D	K	H	T	P	H	T	F	X	U	S	N	U	O	D	K	X	P	N	T	N	G	H
L.	J	X	B	S	K	J	K	Y	H	G	E	U	M	X	N	G	Z	N	G	X	X	H	K	F	Y	D	N	L	U	I
M.	V	A	U	I	J	F	D	H	Z	N	M	N	N	T	K	S	V	U	X	X	K	M	J	N	I	T	J	N	X	X
N.	P	N	T	N	G	H	J	L	D	H	T	P	D	X	I	N	T	J	K	H	T	P	D	U	Y	D	N	H	F	N
P.	F	O	U	G	S	N	G	A	H	G	J	U	G	F	U	O	S	H	T	L	D	I	G	K	H	D	H	F	O	U
Q.	G	S	N	F	H	T	H	J	J	K	H	T	L	N	A	K	Y	D	Y	D	N	L	U	S	S	I	T	K	X	Y
R.	J	N	H	F	N	G	X	D	N	A	H	X	X	I	V	V	U	X	N	F	Y	U	M	N	O	K	P	D	Y	K
S.	T	P	B	X	I	L	D	H	T	H	J	J	K	H	T	L	N	Y	D	N	X	N	U	M	X	N	G	Z	N	G
T.	X	F	N	L	J	H	G	N	F	U	V	N	T	N	F	I	V	H	G	N	F	G	U	I	Y	N	O	G	U	S
V.	S	U	X	L	U	A	Y	U	T	U	G	Y	D	H	T	F	L	N	T	Y	G	H	J	L	D	K	T	H	B	

The uniliteral frequency distribution for this converted text follows. Note that the frequency of each letter is the sum of the five frequencies in the corresponding columns of Fig. 19*b*.

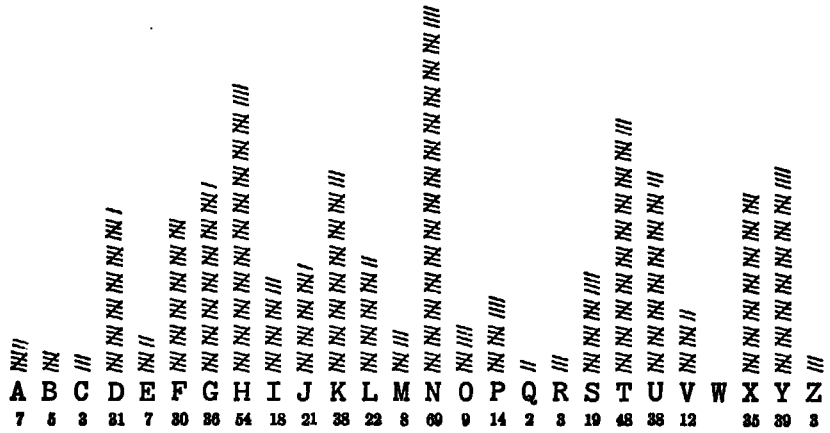


FIGURE 20.

g. The problem having been reduced to monoalphabetic terms, a triliteral frequency distribution can now be made and solution readily attained by simple principles. It yields the following:

JAPAN CONSULTED GERMANY TODAY ON REPORTS THAT THE COMMUNIST INTERNATIONAL WAS BEHIND THE AMAZING SEIZURE OF GENERALISSIMO CHIANG KAI·SHEK IN CHINA. TOKYO ACTED UNDER THE ANTICOMMUNIST ACCORD RECENTLY SIGNED BY JAPAN AND GERMANY. THE PRESS SAID THERE WAS INDISPUTABLE PROOF THAT THE COMINTERN INSTIGATED THE SEIZURE OF GENERAL CHIANG AND SOME OF HIS GENERALS. MILITARY OBSERVERS SAID THE COUP WOULD HAVE BEEN IMPOSSIBLE UNLESS GENERAL CHANG HSUEN LIANG HOTHEADED FORMER WAR LORD OF MANCHURIA HAD FORMED AN ALLIANCE WITH THE COMMUNIST LEADERS HE WAS SUPPOSED TO BE FIGHTING. SUCH AN ALLIANCE THESE OBSERVERS DECLARED OPENED UP A RED ROUTE FROM MOSCOW TO NORTH AND CENTRAL CHINA.

h. The reconstruction of the plain component is now a very simple matter. It is found to be as follows:

H Y D R A U L I C B E F G J K M N O P Q S T V W X Z

Note also, in Fig. 19*b*, the keyword for the message, (HEAVY), the letters being in the columns headed by the letter H.

i. The solution of subsequent messages with different keys can now be reached directly, by a simple modification of the principles explained in Par. 18. This modification consists in using for the completion sequence the *mixed plain component* (now known) instead of the normal alphabet, after the cipher letters have been converted into their plain-component equivalents. Let the student confirm this by experiment.

j. The probable-word method of solution discussed under Paragraph 20 is also applicable here, in case of very short cryptograms. This method presupposes of course, possession of the mixed component and the procedure is essentially the same as that in Par. 20. In the example discussed in the present paragraph, the letter A on the plain component was successively set against the key letters HEAVY; but this is not the only possible procedure.

k. The student should go over carefully the principle of "conversion into monoalphabetic terms" explained in subparagraph *f* above until he thoroughly understands it. Later on he will encounter cases in which this principle is of very great assistance in the cryptanalysis of more complex problems. (Another example will be found under Par. 45.)

l. The principle illustrated in subparagraph *e*, that is, shifting two or more monoalphabetic frequency distributions relatively so as to bring them into proper alignment for amalgamation into a single monoalphabetic distribution, is called *matching*. It is a very important cryptanalytic principle. Note that its practical application consists in sliding one monoalphabetic distribution against the other so as to obtain the best coincidence between the *entire sequence* of crests and troughs of one distribution and the *entire sequence* of crests and troughs of the other distribution. When the best point of coincidence has been found, the two sequences may be amalgamated and *theoretically* the single resultant distribution will also be monoalphabetic in character. The successful application of the principle of matching depends upon several factors. First, the cryptographic situation must be such that matching is a correct cryptographic step. For example, the distributions in figure 19*b* are properly subject to matching because the cipher component in the basic sequences concerned in this problem is the normal sequence, while the plain component is a mixed sequence. But it would be futile to try to match the distributions in figure 9, for in that case the cipher component is a mixed sequence, the plain component is the normal sequence. Hence, no amount of shifting or matching can bring the distributions of

figure 9 into proper superimposition for correct amalgamation. (If the occurrences in the various distributions in figure 9 had been distributed according to the sequence of letters in the mixed component, then matching would be possible; but in order to be able to distribute these occurrences according to the mixed component, the latter has to be *known*—and that is just what is unknown until the problem has been solved.) A second factor involved in successful matching is the number of elements in the two distributions forming the subject of the test. If both of them have very few tallies, there is hardly sufficient information to permit of matching with any degree of assurance that the work is not in vain. If one of them has many tallies, the other only a few, the chances for success are better than before, because the positions of the *blanks* in the two distributions can be used as a guide for their proper superimposition.

m. There are certain mathematical and statistical procedures which can be brought to bear upon the matter of cryptanalytic matching. These will be presented in a later text. However, until the student has studied these mathematical and statistical methods of matching distributions, he will have to rely upon mere ocular examination as a guide to proper superimposition. Obviously, the more data he has in each distribution, the easier is the correct superimposition ascertained by any method.

SECTION VI

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, II

	Paragraph
Further cases to be considered	27
Identical primary mixed components proceeding in the same direction	28
Cryptographing and decryptographing by means of identical primary mixed components	29
Principles of solution	30

27. Further cases to be considered.—*a.* Thus far Cases B (1) and (2), mentioned in Paragraph 6 have been treated. There remains Case B (3), and this case has been further subdivided as follows:

CASE B (3). Both components are mixed sequences.

(a) Components are identical mixed sequences.

(1) Sequences proceed in the same direction. (The secondary alphabet are mixed alphabets.)

(2) Sequences proceed in opposite directions. (The secondary alphabets are reciprocal mixed alphabets.)

(b) Components are different mixed sequences. (The secondary alphabets are mixed alphabets.)

b. The first of the foregoing subcases will now be examined.

28. Identical primary mixed components proceeding in the same direction.—*a.* It is often the case that the mixed components are derived from an easily remembered word or phrase, so that they can be reproduced at any time from memory. Thus, for example, given the key word QUESTIONABLY, the following mixed sequence is derived:

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z

b. By using this sequence as both plain and cipher component, that is, by sliding this sequence against itself, a series of 26 secondary mixed alphabets may be produced. In enciphering a message, sliding strips may be employed with a key word to designate the particular and successive positions in which the strips are to be set, the same as was the case in previous examples of the use of sliding components. The method of designating the positions, however, requires a word or two of comment at this point. In the examples thus far shown, the key letter, as located on the cipher component, was always set opposite A, as located on the plain component; possibly an erroneous impression has been created, *viz*, that this is invariably the rule. This is decidedly not true, as has already been explained in paragraph 7*c*. If it has seemed to be the case that Θ_k always equals A_p, it is only because the text has dealt thus far principally with cases in which the plain component is the normal sequence and its intital letter, which usually constitutes the index for juxtaposing cipher components, is A. It must be emphasized, however, that various conventions may be adopted in this respect; but the most common of them is to employ the initial letter of the plain component as the index letter. That is, the index letter, Θ_i, will be the initial letter of the mixed sequence, in this case, Q. Furthermore, to prevent the possibility of ambiguity it will be stated again that the pair of enciphering equations employed in the ensuing discussion will be the first of the 12 set forth under Par. 7*f*, *viz*, $\Theta_k/_2=\Theta_i/_1$; $\Theta_p/_1=\Theta_c/_2$. In this case the subscript "1" means the plain component, the subscript "2", the cipher component, so that the enciphering equation is the following: $\Theta_k/_c=\Theta_i/_p$; $\Theta_p/_p=\Theta_c/_c$.

c. By setting the two sliding components against each other in the two positions shown below, the cipher alphabets labeled (1) and (2) given by two key letters, A and B, are seen to be different.

KEY LETTER=A

Θ_1 ↓

Plain component..... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
Cipher component..... QUESTIONABLYCDFGHJKMPRVWXZ

↑ Θ_k

Secondary alphabet (1):

Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.....	H	J	P	R	L	V	W	X	D	Z	Q	K	U	G	F	E	A	S	Y	C	B	T	I	O	M	N

KEY LETTER=B

Θ_1 ↓

Plain component..... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
Cipher component..... QUESTIONABLYCDFGHJKMPRVWXZ

↑ Θ_k

Secondary alphabet (2):

Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.....	J	K	R	V	Y	W	X	Z	F	Q	U	M	E	H	G	S	B	T	C	D	L	I	O	N	P	A

d. Very frequently a quadricular or square table is employed by the correspondents, instead of sliding strips, but the results are the same. The cipher square based upon the word QUESTIONABLY is shown in Fig. 21. It will be noted that it does nothing more than set forth the successive positions of the two primary sliding components; the top line of the square is the plain component, the successive horizontal lines below it, the cipher component in its various juxtapositions. The usual method of employing such a square (i. e., corresponding to the enciphering equations $\Theta_{k/e}=\Theta_{1/p}$; $\Theta_{p/p}=\Theta_{e/c}$) is to take as the cipher equivalent of a plain-text letter that letter which lies at the intersection of the vertical column headed by the plain-text letter and the horizontal row begun by the key letter. For example, the cipher equivalent of E_p with keyletter T is the letter O_e; or E_p $(T_k)=O_e$. The method given in paragraph *b*, for determining the cipher equivalents by means of the two sliding strips yields the same results as does the cipher square.

Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z
U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q
E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U
S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E
T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S
I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T
O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I
N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N
B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A
L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B
Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L
C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y
D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C
F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D
G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G
J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H
K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J
M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K
P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M
R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P
V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V
X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W
Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X

FIGURE 21.

29. Cryptographing and decryptographing by identical primary mixed components.—There is nothing of special interest to be noted in connection with the use either of identical mixed components or of an equivalent quadricular table such as that shown in Fig. 21, in enciphering or deciphering a message. The basic principles are the same as in the case of the sliding of one mixed component against the normal, the displacements of the two components being controlled by changeable key words of varying lengths. The components may be changed at will and so on. All this has been demonstrated adequately enough in *Elementary Military Cryptography*, and *Advanced Military Cryptography*.

30. Principles of solution.—*a.* Basically the principles of solution in the case of a cryptogram enciphered by two identical mixed sliding components are the same as in the preceding case. Primary recourse is had to the principles of frequency and repetition of single letters, digraphs, trigraphs, and polygraphs. Once an entering wedge has been forced into the problem, the subsequent steps may consist merely in continuing along the same lines as before, building up the solution bit by bit.

b. Doubtless the question has already arisen in the student's mind as to whether any principles of symmetery of position can be used to assist in the solution and in the reconstruction of the cipher alphabets in cases of the kind under consideration. This phase of the subject will be taken up in the next section and will be treated in a somewhat detailed manner, because the theory and principles involved are of very wide application in cryptanalytics.

SECTION VII

THEORY OF INDIRECT SYMMETRY OF POSITION IN SECONDARY ALPHABETS

Paragraph
Reconstruction of primary components from secondary alphabets 31

31. Reconstruction of primary components from secondary alphabets.—*a.* Note the two secondary alphabets (1) and (2) given in paragraph 28*c*. Externally they show no resemblance or symmetry despite the fact that they were produced from the same primary components. Nevertheless, when the matter is studied with care, a symmetry of position is discoverable. Because it is a hidden or latent phenomenon, it may be termed *latent symmetry of position.* However, in previous texts the phenomenon has been designated as an *indirect symmetry of position* and this terminology has grown into usage, so that a change is perhaps now inadvisable. Indirect symmetry of position is a very interesting and exceedingly useful phenomenon in cryptanalytics.

b. Consider the following secondary alphabet (the one labeled (2) in paragraph 28*c*):

(2)
Plain ..... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher ..... J K R V Y W X Z F Q U M E H G S B T C D L I O N P A

c. Assuming it to be known that this is a secondary alphabet produced by two primary identical mixed components, it is desired to reconstruct the latter. Construct a chain of alternating plain-text and cipher-text equivalents, beginning at any point and continuing until the chain has been completed. Thus, for example, beginning with $A_p=J_c$, $J_p=Q_c$, $Q_p=B_c$, . . ., and dropping out the letters common to successive pairs, there results the sequence A J Q B . . .. By completing the chain the following sequence of letters is established:

A J Q B K U L M E Y P S C R T D V I F W O G X N H Z

d. This sequence consists of 26 letters. *When slid against itself it will produce exactly the same secondary alphabets as do the primary components based upon the word* QUESTIONABLY. To demonstrate that this is the case, compare the secondary alphabets given by the two settings of the externally different components shown below:

Plain component ..... QUESTIONABLYCDFGHJKMPRVWXZQUESTIONABLYCDFGHJKMPRVWXZ
Cipher component ..... QUESTIONABLYCDFGHJKMPRVWXZ

Secondary alphabet (1):

Plain ..... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher ..... J K R V Y W X Z F Q U M E H G S B T C D L I O N P A

Plain component ..... AJQBKULMEYPSCRTDVIFWOGXNHZAJQBKULMEYPSCRTDVIFWOGXNHZ
Cipher component ..... AJQBKULMEYPSCRTDVIFWOGXNHZ

Secondary alphabet (2):

Plain ..... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher ..... J K R V Y W X Z F Q U M E H G S B T C D L I O N P A

e. Since the sequence A J Q B K . . . gives exactly the same equivalents in the secondary alphabets as the sequence Q U E S T . . . gives, the former sequence is cryptographically equivalent to the latter sequence. For this reason the A J Q B K . . . sequence is termed an *equivalent primary component.*[1] If the real or original primary component is a key-word mixed sequence, it is hidden or *latent* within the equivalent primary sequence; but it can be made *patent* by decimation of the equivalent primary component. The procedure is as follows: Find three letters in the equivalent primary component such as are likely to have formed an unbroken sequence in the original primary component, and see if the interval between the first and second is the same as that between the second and third. Such a case is presented by the letters W, X, and Z in the equivalent primary component above. Note the sequence . . . W O G X N H Z . . . ; the distance or interval between the letters W, X, and Z is two letters. Continuing the chain by adding letters two intervals removed, the latent original primary component is made patent. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V

f. It is possible to perform the steps given in *c* and *e* in a combined single operation when the original primary component is a key-word mixed sequence. Starting with any pair of letters (in the cipher component of the secondary alphabet) likely to be sequent in the key-word mixed sequence, such as JK_6 in the secondary alphabet labeled (2), the following chain of digraphs may be set up. Thus, J, K, in the plain component stand over Q, U, respectively, in the cipher component; Q, U, in the plain component stand over B, L, respectively, in the cipher component, and so on. Connecting the pairs in a series, the following results are obtained:

JK → QU → BL → KM → UE → LY → MP → ES → YC → PR → ST → CD → RV →
TI → DF → VW → IO → FG → WX → ON → GH → XZ → NA → HJ → ZQ → AB → JK . . .

These may now be united by means of their common letters:

JK → KM → MP → PR → RV → etc.=J K M P R V W X Z Q U E S T I O N A B L Y C D F G H

The original primary component is thus completely reconstructed.

g. Not all of the 26 secondary alphabets of the series yielded by two sliding primary components may be used to develop a complete equivalent primary component. If examination be made, it will be found that only 13 of these secondary alphabets will yield complete equivalent primary components when the method of reconstruction shown in subparagraph *c* above is followed. For example the following secondary alphabet, which is also derived, from the primary components based upon the word QUESTIONABLY will not yield a complete chain of 26 plain text-cipher-plain text equivalents:

Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher.....	C	D	H	J	O	K	M	P	B	R	V	F	W	Y	L	X	T	Z	N	A	I	Q	U	E	G	S

[1] Such an equivalent component is merely a sequence which has been or can be developed or derived from the original sequence or basic primary component by applying a *decimation* process to the latter; conversely, the original or basic component can be derived from an equivalent component by applying the same sort of process to the equivalent component. By decimation is meant the selection of elements from a sequence according to some fixed interval. For example, the sequence A E I M . . . is derived, by decimation, from the normal alphabet by selecting every fourth letter.

Equivalent primary component:

1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	
A	C	H	P	X	E	O	L	F	K	V	Q	T	A	C	H	. . . (The A C H sequence begins again.)

h. It is seen that only 13 letters of the chain have been established before the sequence begins to repeat itself. It is evident that exactly one-half of the chain has been established. The other half may be established by beginning with a letter not in the first half. Thus:

1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	
B	D	J	R	Z	S	N	Y	G	M	W	U	I	B	D	J	. . . (The B D J sequence begins again.)

i. It is now necessary to distribute the letters of each half-sequence within 26 spaces, to correspond with their placements in a complete alphabet. This can only be done by allowing a constant *odd* number of spaces between the letters of one of the half-sequences. Distributions are therefore made upon the basis of 3, 5, 7, 9, . . . spaces. Select that distribution which most nearly coincides with the distribution to be expected in a key-word component. Thus, for example, with the first half-sequence the distribution selected is the one made by leaving three spaces between the letters. It is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	–	L	–	C	–	F	–	H	–	K	–	P	–	V	–	X	–	Q	–	E	–	T	–	O	–

j. Now interpolate, by the same constant interval (three in this case), the letters of the other half-sequence. Noting that the group F – H appears in the foregoing distribution, it is apparent that G of the second half-sequence should be inserted between F and H. The letter which immediately follows G in the second half-sequence, *viz*, M, is next inserted in the position three spaces to the right of G, and so on, until the interpolation has been completed. This yields the original primary component, which is as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	B	L	Y	C	D	F	G	H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N

k. Another method of handling cases such as the foregoing is indicated in subparagraph *f* By extending the principles set forth in that subparagraph, one may reconstruct the following chain of 13 pairs from the secondary alphabet given in subparagraph *g:*

1		2		3		4		5		6		7		8		9		10		11		12		13		1	
CD	→	HJ	→	PR	→	XZ	→	ES	→	ON	→	LY	→	FG	→	KM	→	VW	→	QU	→	TI	→	AB	→	CD	. . .

Now find, in the foregoing chain, two pairs likely to be sequent, for example HJ and KM and count the interval between them in the chain. It is 7 (counting by pairs). If this decimation interval is now applied to the chain of pairs, the following is established:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	J	K	M	P	R	V	W	X	Z	Q	U	E	S	T	I	O	N	A	B	L	Y	C	D	F	G

l. The reason why a complete chain of 26 letters cannot be constructed from the secondary alphabet given under subparagraph *g* is that it represents a case in which two primary components of 26 letters were slid an *even* number of intervals apart. (This will be explained in further detail in subparagraph *r* below.) There are in all 12 such cases, none of which will admit of the construction of a complete chain of 26 letters. In addition, there is one case wherein, despite the fact that the primary components are an *odd* number of intervals apart, the secondary alphabet cannot be made to yield a complete chain of 26 letters for an equivalent primary component. This is the case in which the displacement is 13 intervals. Note the secondary alphabet based upon the primary components below (which are the same as those shown in subparagraph *d*):

PRIMARY COMPONENTS

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z
D F G H J K M P R V W X Z Q U E S T I O N A B L Y C

SECONDARY ALPHABET

Plain..... A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Cipher..... R V Z Q G U E S K T I W O P M N D A H J F B L Y X C

m. If an attempt is made to construct a chain of letters from this secondary alphabet alone, no progress can be made because the alphabet is completely reciprocal. However, the cryptanalyst need not at all be baffled by this case. The attack will follow along the lines shown below in subparagraphs *n* and *o*.

n. If the original primary component is a key-word mixed sequence, the cryptanalyst may reconstruct it by attempting to "dovetail" the 13 reciprocal pairs (AR, BV, CZ, DQ, EG, FU, HS, IK, JT, LW, MO, NP, and XY) into one sequence. The members of these pairs are all 13 intervals apart. Thus:

0	1	2	3	4	5	6	7	8	9	10	11	12	13
A	.	.	.	.	.	.	.	.	.	.	.	.	R
B	.	.	.	.	.	.	.	.	.	.	.	.	V
C	.	.	.	.	.	.	.	.	.	.	.	.	Z
D	.	.	.	.	.	.	.	.	.	.	.	.	Q
E	.	.	.	.	.	.	.	.	.	.	.	.	G
F	.	.	.	.	.	.	.	.	.	.	.	.	U
H	.	.	.	.	.	.	.	.	.	.	.	.	S
I	.	.	.	.	.	.	.	.	.	.	.	.	K
J	.	.	.	.	.	.	.	.	.	.	.	.	T
L	.	.	.	.	.	.	.	.	.	.	.	.	W
M	.	.	.	.	.	.	.	.	.	.	.	.	O
N	.	.	.	.	.	.	.	.	.	.	.	.	P
X	.	.	.	.	.	.	.	.	.	.	.	.	Y

FIGURE 22.

Write out the series of numbers from 1 to 26 and insert as many pairs into position as possible. being guided by considerations of probable partial sequences in the key-word mixed sequence, Thus:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	B	C	D	.	.	.	.	.	.	.	.	.	R	V	Z	Q

It begins to look as though the key-word commences with the letter Q, in which case it should be followed by U. This means that the next pair to be inserted is FU. Thus:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	B	C	D	F	.	.	.	.	.	.	.	.	R	V	Z	Q	U

The sequence A B C D F means that E is in the key. Perhaps the sequence is A B C D F G H. Upon trial, using the pairs EG and HS, the following placements are obtained:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	B	C	D	F	G	H	.	.	.	.	.	.	R	V	Z	Q	U	E	S

This suggests the word QUEST or QUESTION. The pair JT is added:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	B	C	D	F	G	H	J	.	.	.	.	.	R	V	Z	Q	U	E	S	T

The sequence G H J suggests G H J K, which places an I after T. Enough of the process has been shown to make the steps clear.

o. Another method of circumventing the difficulties introduced by the 14th secondary alphabet (displacement interval, 13) is to use it in conjunction with another secondary alphabet which is produced by an even-interval displacement. For example, suppose the following two secondary alphabets are available.[1]

Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	R	V	Z	Q	G	U	E	S	K	T	I	W	O	P	M	N	D	A	H	J	F	B	L	Y	X	C
2	X	Z	E	S	K	T	I	O	R	N	A	Q	B	W	V	L	H	Y	M	P	J	C	D	F	U	G

FIGURE 23.

The first of these secondaries is the 13-interval secondary; the second is one of the even-interval secondaries, from which only half-chain sequences can be constructed. But if the construction be based upon the two sequences, 1 and 2 in the foregoing diagram, the following is obtained:

R X U T N L D H M V Z E I A Y F J P W Q S O B C G K

This is a complete equivalent primary component. The original key-word mixed component can be recovered from it by decimation based upon the 9th interval:

R V W X Z Q U E S T I O N A B L Y C D F G H J K M P

p. (1) When the primary components are identical mixed sequences proceeding in *opposite* directions, all the secondary alphabets will be reciprocal alphabets. Reconstruction of the primary component can be accomplished by the procedure indicated under subparagraph *o* above. Note the following three reciprocal secondary alphabets:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	P	M	H	G	Q	F	D	C	W	Y	L	K	B	R	V	A	E	N	Z	X	U	O	I	T	J	S
2	W	V	M	K	S	J	H	G	Q	F	D	R	C	X	Z	Y	I	L	E	U	T	B	A	N	P	O
3	T	S	S	Z	L	X	W	V	N	R	P	E	M	I	O	K	C	J	B	A	Y	H	G	F	U	D

FIGURE 24.

(2) Using lines 1 and 2, the following chain can be constructed (equivalent primary component):

P W Q S O B C G K R X U T N L D H M V Z E I A T F J

[1] The method of writing down the secondaries shown in figure 23 will hereafter be followed in all cases when alphabet reconstruction skeletons are necessary. The top line will be understood to be the plain component; it is common to all the secondary alphabets, and is set off from the cipher components by the heavy black line. This top line of letters will be designated by the digit Ø, and will be referred to as "the zero line" in the diagram. The successive lines of letters, which occupy the space below the zero line and which contain the various cipher components of the several secondary alphabets, will be numbered serially. These numbers may then be used as reference numbers for designating the horizontal lines in the diagram. The numbers standing above the letters may be used as reference numbers for the vertical columns in the diagram. Hence, any letter in the reconstruction skeleton may be designated by coordinates, giving the horizontal or X coordinate first. Thus, D (2–11) means the letter D standing in line 2, Column 11.

Or, using lines 2 and 3:

W T Y K Z O D P U A G V S L J X I C M Q N F R E B H

The original key-word mixed primary component (based on the word QUESTIONABLY) can be recovered from either of the two foregoing equivalent primary components. But if lines 1 and 3 are used, only half-chains can be constructed:

P T F X A K E C V O H Q L and M S D W N J U Y R I G Z B

This is because 1 and 3 are both odd-interval secondary alphabets, whereas 2 is an even-interval secondary. It may be added that odd-interval secondaries are characterized by having two cases in which a plain-text letter is enciphered by itself; that is, Θ_p is identical with Θ_c. This phrase "identical with" will be represented by the symbol $\equiv$; the phrase "not identical with" will be represented by the symbol $\not\equiv$. (Note that in secondary alphabet number 1 above, $F_p \equiv F_c$ and $U_p \equiv U_c$; in secondary alphabet number 3 above, $M_p \equiv M_c$ and $O_p \equiv O_c$). This characteristic will enable the cryptanalyst to select at once the proper two secondaries to work with in case several are available; one should show two cases where $\Theta_p \equiv \Theta_c$; the other should show none.

q. (1) When the primary components are different mixed sequences, their reconstruction from secondary cipher alphabets follows along the same lines as set forth above, under *b* to *j*, inclusive, with the exception that the selection of letters for building up the chain of equivalents for the primary cipher component is restricted to those below the zero line in the reconstruction skeleton. Having reconstructed the primary cipher component, the plain component can be readily reconstructed. This will become clear if the student will study the following example.

Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	T	V	A	B	U	L	I	Q	X	Y	C	W	S	N	D	P	F	E	Z	G	R	H	J	K	M	O
2	Z	J	S	T	V	I	Q	R	M	O	N	K	X	E	A	G	B	W	P	L	H	Y	C	D	F	U

FIGURE 25.

(2) Using only lines 1 and 2, the following chain is constructed:

T Z P G L I Q R H Y O U V J C N E W K D A S X M F B

This is an equivalent primary cipher component. By finding the values of the successive letters of this chain in terms of the plain component of secondary alphabet number 1 (the zero line), the following is obtained:

T Z P G L I Q R H Y O U V J C N E W K D A S X M F B
A S P T F G H U V J Z E B W K N R L X O C M I Y Q D

The sequence A S P T . . . is an equivalent primary plain component. The original key-word mixed components may be recovered from each of the equivalent primary components. That for the primary plain component is based upon the key PUBLISHERS MAGAZINE; that for the primary cipher component is based upon the key QUESTIONABLY.

(3) Another method of accomplishing the process indicated above can be illustrated graphically by the following two chains, based upon the two secondary alphabets set forth in subparagraph *q* (1):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ø.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1.....	T	V	A	B	U	L	I	Q	X	Y	C	W	S	N	D	P	F	E	Z	G	R	H	J	K	M	O
2.....	Z	J	S	T	V	I	Q	R	M	O	N	K	X	E	A	G	B	W	P	L	H	Y	C	D	F	U

Col. 1.		Col. 2.				
A (Ø–1)	→	T (1–1);	→	T (2–4)	→	D (Ø–4); →
D (Ø–4)	→	B (1–4);	→	B (2–17)	→	Q (Ø–17); →
Q (Ø–17)	→	F (1–17);	→	F (2–25)	→	Y (Ø–25); →
Y (Ø–25)	→	M (1–25);	→	M (2–9)	→	I (Ø–9); →
I (Ø–9)	→	X (1–9);	→	X (2–13)	→	M (Ø–13); →
M (Ø–13)	→	S (1–13);	→	S (2–3)	→	C (Ø–3); →
etc.		etc.				

FIGURE 26.

(4) By joining the letters in Column 1, the following chain is obtained: A D Q Y I M, etc. If this be examined, it will be found to be an equivalent primary of the sequence based upon PUBLISHERS MAGAZINE. By joining the letters in Column 2, the following chain is obtained: T B F M X S. This is an equivalent primary of the sequence based upon QUESTIONABLY.

r. A final word concerning the reconstruction of primary components in general may be added. It has been seen that in the case of a 26-element component sliding against itself (both components proceeding in the same direction), it is only the secondary alphabets resulting from odd-interval displacements of the primary components which permit of reconstructing a single 26-letter chain of equivalents. This is true except for the 13th interval displacement, which even though an odd number, still acts like an even number displacement in that no complete chain of equivalents can be established from the secondary alphabet. This exception gives the clue to the basic reason for this phenomenon: it is that the number 26 has two factors, 2 and 13, which enter into the picture. With the exception of displacement-interval 1, *any displacement interval which is a sub-multiple of, or has a factor in common with the number of letters in the primary sequence will yield a secondary alphabet from which no complete chain of 26 equivalents can be derived for the construction of a complete equivalent primary component.* This general rule is applicable only to components which progress in the same direction; if they progress in opposite directions, all the secondary alphabets are reciprocal alphabets and they behave exactly like the reciprocal secondaries resulting from the 13-interval displacement of two 26-letter identical components progressing in the same direction.

s. The foregoing remarks give rise to the following observations based upon the general rule pointed out above. Whether or not a complete equivalent primary component is derivable by decimation from an original primary component (and if not, the lengths and numbers of chains of letters, or incomplete components, that can be constructed in attempts to derive such equivalent components) will depend upon the number of letters in the original primary component and the specific decimation interval selected. For example, in a 26-letter original primary component, decimation interval 5 will yield a complete equivalent primary component of 26 letters, whereas decimation intervals 4 or 8 will yield 2 chains of 13 letters each. In a 24-letter component, decimation interval 5 will also yield a complete equivalent primary component (of 24 letters), but decimation interval 4 will yield 6 chains of 4 letters each, and decimation interval 8 will

yield 3 chains of 8 letters each. It also follows that in the case of an original primary component in which the total number of characters is a prime number, *all* decimation intervals will yield complete equivalent primary components. The following table has been drawn up in the light of these observations, for original primary sequences from 16 to 32 elements. (All prime-number sequences have been omitted.) In this table, the column at the extreme left gives the various decimation intervals, omitting in each case the first interval, which merely gives the original primary sequence, and the last interval, which merely gives the original sequence reversed. The top line of the table gives the various lengths of original primary sequences from 32 down to 16. (The student should bear in mind that sequences containing characters in addition to the letters of the alphabet may be encountered; he can add to this table when he is interested in sequences of more than 32 characters.) The numbers within the table then show, for each combination of decimation interval and length of, original sequence, the lengths of the chains of characters that can be constructed. (The student may note the symmetry in each column.) The bottom line shows the total number of complete equivalent primary components which can be derived for each different length of original component.

Decimation interval	Number of characters in original primary component											
	32	30	28	27	26	25	24	22	21	20	18	16
2	16	15	14	27	13	25	12	11	21	10	9	8
3	32	10	28	9	26	25	8	22	7	20	6	16
4	8	15	7	27	13	25	6	11	21	5	9	4
5	32	6	28	27	26	5	24	22	21	4	18	16
6	16	5	14	9	13	25	4	11	7	10	3	8
7	32	30	4	27	26	25	24	22	3	20	18	16
8	4	15	7	27	13	25	6	11	21	5	9	2
9	32	10	28	9	26	25	8	22	7	20	2	16
10	16	3	14	27	13	5	12	11	21	2	9	8
11	32	30	28	27	26	25	24	2	21	20	18	16
12	8	5	7	9	13	25	2	11	7	5	3	4
13	32	30	28	27	2	25	24	22	21	20	18	16
14	16	15	2	27	13	25	12	11	3	10	9	8
15	32	2	28	9	26	5	8	22	7	4	6	
16	2	15	7	27	13	25	6	11	21	5	9	
17	32	30	28	27	26	25	24	22	21	20		
18	16	5	14	9	13	25	4	11	7	10		
19	32	30	28	27	26	25	24	22	21			
20	8	3	7	27	13	5	6	11				
21	32	10	4	9	26	25	8					
22	16	15	14	27	13	25	12					
23	32	30	28	27	26	25						
24	4	5	7	9	13							
25	32	6	28	27								
26	16	15	14									
27	32	10										
28	8	15										
29	32											
30	16											
Total number of complete sequences	14	6	10	16	10	18	16	8	10	6	4	6

SECTION VIII

APPLICATION OF PRINCIPLES OF INDIRECT SYMMETRY OF POSITION

	Paragraph
Applying the principles to a specific example	32
The cryptogram employed in the exposition	33
Fundamental theory	34
Application of principles	35
General remarks	36

32. Applying the principles to a specific example.—*a.* The preceding section, with the many details covered, now forms a sufficient base for proceeding with an exposition of how the principles of indirect symmetry of position can be applied very early in the solution of a polyalphabetic substitution cipher in which sliding primary components were employed to produce the secondary cipher alphabets for the enciphering of the cryptogram.

b. The case described below will serve not only to explain the method of applying these principles but will at the same time show how their application greatly facilitates the solution of a single, rather difficult, polyalphabetic substitution cipher. It is realized, of course, that the cryptogram could be solved by the usual methods of frequency and long, patient experimentation. However, the method to be described was actually applied and very materially reduced the amount of time and labor that would otherwise have been required for solution.

33. The cryptogram employed in the exposition.—*a.* The problem that will be used in this exposition involves an actual cryptogram submitted for solution in connection with a cipher device having two concentric disks upon which the same random mixed alphabet appears, both alphabets progressing in the same direction. This was obtained from a study of the descriptive circular accompanying the cryptogram. By the usual process of factoring, it was determined that the cryptogram involved 10 alphabets. The message as arranged according to its period is shown in Figure 27, in which all repetitions of two or more letters are indicated.

b. The triliteral frequency distributions are given in Figure 28. It will be seen that on account of the brevity of the message, considering the number of alphabets involved, the frequency distributions do not yield many clues. By a very careful study of the repetitions, tentative individual determinations of values of cipher letters, as illustrated in Figures 29, 30, 31, and 32, were made. These are given in sequence and in detail in order to show that there is nothing artificial or arbitrary in the preliminary stages of analysis here set forth.

The Cryptogram

(Repetitions underlined)

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
B	G	B	Z	D	P	F	B	O	U	O
C	G	R	F	T	Z	M	Q	M	A	V
D	K	Z	U	G	D	Y	F	T	R	W
E	G	J	X	N	L	W	Y	O	U	X
F	I	K	W	E	P	Q	Z	O	K	Z
G	P	R	X	D	W	L	Z	I	C	W
H	G	K	Q	H	O	L	O	D	V	M
I	G	O	X	S	N	Z	H	A	S	E
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
N	G	S	X	E	R	O	Z	J	S	E
O	G	V	Q	W	E	J	M	K	G	H

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
S	C	F	B	S	C	V	X	C	H	Q
T	Z	T	Z	S	D	M	X	W	C	M
U	R	K	U	H	E	Q	E	D	G	X
V	F	K	V	H	P	J	J	K	J	Y
W	Y	Q	D	P	C	J	X	L	L	L
X	G	H	X	E	R	O	Q	P	S	E
Y	G	K	B	W	T	L	F	D	U	Z
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
BB	H	S	P	O	P	N	M	D	L	M
CC	G	C	K	W	D	V	B	L	S	E
DD	G	S	U	G	D	P	O	T	H	X

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
FF	L	F	U	Y	D	T	Z	V	H	Q
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
PP	U	K	W	P	E	F	X	E	N	F
QQ	C	C	U	G	D	W	P	E	U	H
RR	Y	B	W	E	W	V	M	D	W	J
SS	R	Z	X							

Figure 27

152018—38——5

Triliteral Frequency Distributions

I

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	EB	FF			XK	YB	ES	XK	ZC	VZ	WQ			ZC	ZR	DC	HC	HR		MK		-F		YQ	QT
	HZ	FC				OR	NH		VQ	ZL	JF						MK							NT	QG
	XK					WJ	ZO		QJ								JZ							NU	
	WG					WK																		HB	
	QK					MO																			
						ES																			
						EV																			
						LH																			
						EK																			
						MC																			
						ES																			

II

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	GZ	QB			WU	ZW	GX		GX	IW	KB			GX		LZ	GF	GX	ZZ	YX	GQ				KU
	BJ	JQ			CB	BB	HV		JU	GQ				HZ		YD	PX	HP	YX						BZ
	YW	RV			LU					RU						JW	SQ	GU							RX
		OD								FV															
		GK								GB															
		CU								BD															
										BX															
										UW															

III

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	CZ		QP		RT				BI	CW					SO	KH				FP	CO	KE	JN		BD
	FS		CH													CR				ZG	KH	GN	RD		QA
	KW		KZ													RE				KH	HL	QO	OS		ZC
	LP															VW				SG		KP	SE		TS
	GW																			FY		BE	HE		OO
																				JG			TC		
																				CG			KD		
																							UO		
																							Z-		

IV

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ZA		ZK	ZP	WP		UD	QO	JP			VA		XL	VP	UC		QQ	XN	FZ			QE		UD	BE
		XD	XW	QW		UD	UE						WK	PP	DC			BC				BT			DF
			XS	XR		UD	VP							WO	BC			ZD				KD			
				XR		UD	DW							XP	WE							BW			
				WW										ZW											

V

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
AA		PF	GY	ZX	ZM					CQ	NW		SZ	HL	DF	RF	EO	DO	WL			DL			TM
LQ		SV	SM	WJ						NX				OT	EQ		EO					EM			
		PJ	WV	HQ											IQ							HM			
		PJ	GP	PF											ON							WO			
			YT												HJ							OM			
			GP												ON							EV			
			GW												OP										
			GW																						

VI

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
AM					CO				EM		WZ	ZQ	PB	RZ	DO	PZ			DZ		CX	LY	EQ	DF	NH
					PB				PJ		OO	WL	PM	RQ	DM	PF			OT		DB	DQ	KJ		
					QV				CX		TF	DX		WQ	PY	KO					WM	DP			
					EX				CO			WZ		SZ		EE										
												FT				AQ									
												WX													

VII

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	FO			QD	YT		ZA		JK		MN	JK		FC	WE	MM			MG		FM		VC	WO	QO
	NL				QJ				XT			AD		LD		XT			TN				MW	PO	LI
	VL				LD							ND		QI		OP							JL		OJ
												PV		JT		OR							MC		MT
												VD		PT		QV							FE		TV
																WR									OR

VIII

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
HS		OJ	OV	XN		TQ		ZC	FH	MG	BC	QA	LA	BU	QS		QG		FR		ZH	XC			
		XH	MC	PU				OK	ZS	JJ	XL	VL	TV	YU			ZS		QX		ML				
		XG	EG								BS			ZK			QV		ZU		QA				
			FU											YX					OX						
			ML																OH						
			MY																JR						

IX

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
MV		IW				KH	JD		CY	OZ	MH		EF			GJ	TW	AE		OO	DM		TZ	DJ	
NE		LW				DX	CQ		KY	IF	LL						TN	JE		OX	NQ		TE		
VV		DH				RN	TX				DM							PE		DZ	RM		OZ		
		WM				CQ	VQ				VW							LE		TZ					
																		RN		EH					

X

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	Z	Y	Z
			HQ	SB	KC		LS		QL		LG	VG	RY	UG		HZ					AK	RG	UI	JG	KP
				AG	NC		GR		YR			CR	GH			HZ					AJ	CG	GF	JY	XJ
				SG			CB					LG	SY			VB						CL	HB		UO
				SG			UY					VU				GJ						LB			UK
				XH																					XH
				SG																					

FIGURE 28.

Initial Values From Assumptions

$G_c = E_p$; $K_c = E_p$; $X_c = E_p$; and $D_c = E_p$, from frequency considerations.

UGD=THE; PCJ=THE; and SEG=THE, from study of repetitions.

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
			T	T	H					
B	G	B	Z	D	P	F	B	O	U	O
	E									
C	G	R	F	T	Z	M	Q	M	A	V
	E									
D	K	Z	U	G	D	Y	F	T	R	W
			T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E								
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E								
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
										H
N	G	S	X	E	R	O	Z	J	S	E
	E		E						T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
S	C	F	B	S	C	V	X	C	H	Q
					H					
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E				E				
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E						T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
				T	H	E				H
BB	H	S	P	O	P	N	M	D	L	M
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E		T	H	E					

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
			T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E						T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T						
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
SS	R	Z	X							
			E							

Figure 29.

Additional Values from Assumptions (I)

Refer to line DD in Figure 29; S_e^{2} assumed to be N_p.

Refer to line M in figure 29; A_e^{9} assumed to be W_p.

Then in lines C–D, $\overset{9}{A}\ \overset{10}{V}\ \overset{1}{K}\ \overset{2}{Z}\ \overset{3}{U}\ \overset{4}{G}\ \overset{5}{D}$ is assumed to be WITH THE.

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
			T	T	H					
B	G	B	Z	D	P	F	B	O	U	O
	E									
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E								
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E								
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
M	S	R	Q	E	W	M	L	N	A	E
									W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E						T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H								
S	C	F	B	S	C	V	X	C	H	Q
					H					
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E				E				
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E						T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E				
BB	H	S	P	O	P	N	M	D	L	M
		N								
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	N	T	H	E					

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
			T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E						T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T						
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
SS	R	Z	X							
		H	E							

Figure 30.

Additional Values from Assumptions (II)

Refer to Figure 30, line A; W F U P C F O C J Y; assume to be BUT THOUGH.
1 2 3 4 5 6 7 8 9 10
– – T T H – – – – –

Refer to Figure 30, lines N and X, where repetition X E R O occurs; assume EACH
3 4 5 6
E – – –

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
	B	U	T	T	H	O	U	G	H	
B	G	B	Z	D	P	F	B	O	U	O
	E					O				
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E		A						
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E					U			
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	Q	R	Q	F	V	M	L	H
						O				
M	S	R	Q	E	W	M	L	N	A	E
				A					W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E	A	C	H			T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H					U			
S	C	F	B	S	C	V	X	C	H	Q
		U			H			G		
T	Z	T	Z	S	D	M	X	W	C	M
					E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E					E		H	
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E	A	C	H			T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E	U			H
BB	H	S	P	O	P	N	M	D	L	M
		N								
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	N	T	H	E		U			

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
		U	T		E					
GG	Z	G	W	N	K	X	J	T	R	N
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
						H				
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
LL	B	K	X	D	S	O	Z	R	S	N
		E	E			H			T	
MM	Y	U	X	O	P	P	Y	O	X	Z
NN	H	O	Z	O	W	M	X	C	G	Q
								G		
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T		O				
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
				A						
SS	R	Z	X							
		H	E							

Figure 31.

Additional Values From Assumptions (III)

$^{4\,5\,6}$
OPN—assume ING from repetition and frequency.

$^{9\,0\,1}$
HQZ—assume ING from repetition and frequency.

	1	2	3	4	5	6	7	8	9	10
A	W	F	U	P	C	F	O	C	J	Y
	B	U	T	T	H	O	U	G	H	
B	G	B	Z	D	P	F	B	O	U	O
	E				N	O				
C	G	R	F	T	Z	M	Q	M	A	V
	E								W	I
D	K	Z	U	G	D	Y	F	T	R	W
	T	H	T	H	E					
E	G	J	X	N	L	W	Y	O	U	X
	E		E							
F	I	K	W	E	P	Q	Z	O	K	Z
		E		A	N					
G	P	R	X	D	W	L	Z	I	C	W
			E							
H	G	K	Q	H	O	L	O	D	V	M
	E	E					U			
I	G	O	X	S	N	Z	H	A	S	E
	E		E						T	H
J	B	B	J	I	P	Q	F	J	H	D
					N				I	
K	Q	C	B	Z	E	X	Q	T	X	Z
L	J	C	O	R	Q	F	V	M	L	H
						O				
M	S	R	Q	E	W	M	L	N	A	E
				A					W	H
N	G	S	X	E	R	O	Z	J	S	E
	E	N	E	A	C	H			T	H
O	G	V	Q	W	E	J	M	K	G	H
	E					E				

	1	2	3	4	5	6	7	8	9	10
P	R	C	V	O	P	N	B	L	C	W
				I	N	G				
Q	L	Q	Z	A	A	A	M	D	C	H
R	B	Z	Z	C	K	Q	O	I	K	F
		H					U			
S	C	F	B	S	C	V	X	C	H	Q
		U			H			G	I	N
T	Z	T	Z	S	D	M	X	W	C	M
	G				E					
U	R	K	U	H	E	Q	E	D	G	X
		E	T							
V	F	K	V	H	P	J	J	K	J	Y
		E			N	E			H	
W	Y	Q	D	P	C	J	X	L	L	L
				T	H	E				
X	G	H	X	E	R	O	Q	P	S	E
	E		E	A	C	H			T	H
Y	G	K	B	W	T	L	F	D	U	Z
	E	E								
Z	O	C	D	H	W	M	Z	T	U	Z
AA	K	L	B	P	C	J	O	T	X	E
	T			T	H	E	U			H
BB	H	S	P	O	P	N	M	D	L	M
		N		I	N	G				
CC	G	C	K	W	D	V	B	L	S	E
	E				E				T	H
DD	G	S	U	G	D	P	O	T	H	X
	E	H	T	H	E		U		I	

	1	2	3	4	5	6	7	8	9	10
EE	B	K	D	Z	F	M	T	G	Q	J
		E								
FF	L	F	U	Y	D	T	Z	V	H	Q
		U	T		E				I	N
GG	Z	G	W	N	K	X	J	T	R	N
	G									
HH	Y	T	X	C	D	P	M	V	L	W
			E		E					
II	B	G	B	W	W	O	Q	R	G	N
						H				
JJ	H	H	V	L	A	Q	Q	V	A	V
									W	I
KK	J	Q	W	O	O	T	T	N	V	Q
				I						N
LL	B	K	X	D	S	O	Z	R	S	N
		E	E			H			T	
MM	Y	U	X	O	P	P	Y	O	X	Z
				I	N					
NN	H	O	Z	O	W	M	X	C	G	Q
				I				G		N
OO	J	J	U	G	D	W	Q	R	V	M
			T	H	E					
PP	U	K	W	P	E	F	X	E	N	F
		E		T		O				
QQ	C	C	U	G	D	W	P	E	U	H
			T	H	E					
RR	Y	B	W	E	W	V	M	D	Y	J
				A						
SS	R	Z	X							
		H	E							

Figure 32.

c. From the initial and subsequent tentative identifications shown in Figures 29, 30, 31, and 32, the values obtained were arranged in the form of the secondary alphabets in a reconstruction skeleton, shown in Figure 33.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1		W			G		Z													K						
2					K			Z						S							F					
3					X															U						
4	E							G	O											P						
5			R		D			C						P												
6					J		N	O							F											
7																					O					
8							C																			
9								J	H											S			A			
10								E	V																	

FIGURE 33.

34. Fundamental theory.—*a.* In paragraph 31, methods of reconstructing primary components from secondary alphabets were given in detail. It is necessary that those methods be fully understood before the following steps be studied. It was there shown that the primary component can be one of a series of equivalent primary sequences, all of which will give exactly similar results so far as the secondary alphabets and the cryptographic text are concerned. It is not necessary that the identical or original primary component employed in the cryptographing be reconstructed; any equivalent primary sequence will serve. The whole question is one of establishing a sequence of letters the interval between which is either identical with that in the original primary component or else is an exact constant multiple of the interval separating the letters in the original primary component. For example, suppose K P X N Q forms a sequence in the original primary component. Here the interval between K and P, and P and X, X and N, N and Q is one; in an equivalent primary component, say the sequence K . . P . . X . . N . . Q, the interval between K and P is three, that between P and X also three, and so on; and the two sequences will yield the same secondary alphabets. So long as the interval between K and P, P and X, X and N, N and Q, . . . , is a constant one, the sequence will be cryptographically equivalent to the original primary sequence and will yield the same secondary alphabets as do those of the original primary sequence. However, in the case of a 26-letter component, it is necessary that this interval be an odd number other than 13, as these are the only cases which will yield one unbroken sequence of 26 letters. Suppose a secondary alphabet to be as follows:

(1)	Plain.....	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Cipher.....														X		K	N							P		

It can be said that the primary component contains the following sequences:

XN KP NQ PX

These, when united by means of their common letters, yield K P X N Q.

Suppose also the following secondary alphabet is at hand:

(2)

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cipher														P			X							K		N

Here the sequences PN, XQ, KX, and NZ can be obtained, which when united yield the two sequences KXQ and PNZ.

By a comparison of the sequences K P X N Q, K X Q, and P N Z, one can establish the following:

K P X N Q
K . X . Q
P . N . Z

It follows that one can now add the letter Z to the sequence, making it K P X N Q Z.

b. The reconstruction of a primary component from one of the secondary alphabets by the process given in paragraph 31 requires a complete or nearly complete secondary alphabet. This is at hand only *after* a cryptogram has been completely solved. But if one could employ several very scant or skeletonized secondary alphabets simultaneously with the analysis of the cryptogram, one could then possibly build up a primary component from fewer data and thus solve the cryptogram much more rapidly than would otherwise be possible.

c. Suppose only the cipher components of the two secondary alphabets (1) and (2) given above be placed into juxtaposition. Thus:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
(1)	.	.	.	.	.	.	.	.	.	.	.	.	.	X	.	K	N	.	.	.	.	.	.	P	.	.
(2)	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.	.	X	.	.	.	.	.	.	K	.	N

The sequences PX, XN, and KP are given by juxtaposition. These, when united, yield KPXN as part of the primary sequence. It follows, therefore, that *one can employ the cipher components of secondary alphabets as sources of independent data* to assist in building up the primary sequences. The usefulness of this point will become clearer subsequently.

35. Application of principles.—*a.* Refer now to the reconstruction skeleton shown in Figure 33. Hereafter, in order to avoid all ambiguity and for ease in reference, the position of a letter in Figure 33 will be indicated as stated in footnote 1, page 56. Thus, N (6–7) refers to the letter N in line 6 and in column 7 of Figure 33.

b. (1) Now, consider the following pairs of letters:

E (Ø–5) J (6–5)
G (Ø–7) N (6–7)
H (Ø–8) O (6–8)
O (Ø–15) F (6–15)

HO, OF=HOF

(One is able to use the line marked zero in Figure 33 since this is a mixed sequence sliding against *itself*.)

(2) The immediate results of this set of values will now be given. Having HOF as a sequence, with EJ as belonging to the same displacement interval, suppose HOF and EJ are placed into juxtaposition as portions of sliding components. Thus:

Plain..... . . . H O F . . .
Cipher.... . . . E J . . . .

When $H_p=E_c$, then $O_p=J_c$.

(3) Refer now to alphabet 10, Figure 33, where it is seen that $H_p=E_c$. *The derived value,* $O_p=J_c$, *can immediately be inserted in the same alphabet* and substituted in the cryptogram.

(4) The student may possibly get a clearer idea of the principles involved if he will regard the matter as though he were dealing with arithmetical proportion. For instance, given any three terms in the proportion 2:8=4:16, the 4th term can easily be found. Furthermore, given the pair of values on the left-hand side of the equation, one may find numerous pairs of values which may be inserted in the right-hand side, or vice versa. For instance, 2:8=4:16 is the same as 2:8=5:20, or 9:36=4:16, and so on. An illustration of each of these principles will now be given, reference being made to Figure 33. As an example of the first principle, note that E (Ø–5):H (Ø–8)=J (6–5):O (6–8). Now find E (10–8):H (Ø–8)=? (10–15):O (Ø–15). It is clear that J may be inserted as the 3d term in this proportion, thus giving the important new value, $\overset{10}{O_p}=J_c$, which is exactly what was obtained directly above, by means of the partial sliding components. As an example of the second principle, note the following pairs:

E (Ø–5) H (Ø–8)
K (2–5) Z (2–8)
D (5–5) C (5–8)
J (6–5) O (6–8)

These additional pairs are also noted:

K (1–20) Z (1–7)
T (Ø–20) G (Ø–7)

Therefore, E:H=K:Z=D:C=J:O=T:G, and T may be inserted in position (4–5).

c. (1) Again, GN belongs to the same set of displacement-interval values as do EJ and HOF. Hence, by superimposition:

Plain..... . . . H O F . . .
Cipher.... . . . G N . . . .

(2) Referring to alphabet 4, when $H_p=G_c$, then $O_p=N_c$. Therefore, the letter N can be inserted in position (4–15) in Figure 33, and the value $\overset{4}{N_c}=O_p$ can be substituted in the cryptogram.

(3) Furthermore, note the corroboration found from this particular superimposition:

H (Ø–8) G (Ø–7)
O (6–8) N (6–7)

This checks up the value in alphabet 6, $G_p=N_c$.

d. (1) Again superimpose HOF and GN:

. . . H O F . . .
. . . . G N . . .

(2) Note this corroboration:

O (6–8) G (4–8)
F (6–15) N (4–15)

which has just been inserted in Figure 33, as stated above.

e. (1) Again using HOF and EJ, but in a different superimposition:

. . . H O F . .
. . E J . . . .

(2) Refer now to H (9–9), J (9–8). Directly under these letters is found V (10–9), E (10–8).

Therefore, the V can be added immediately before H O F, making the sequence V H O F.

f. (1) Now take V H O F and juxtapose it with E J, thus:

. . . V H O F . . .
. . . E J . . .

(2) Refer now to Figure 33, and find the following:

V (10–9)	E (10–8)
H (9–9)	J (9–8)
O (4–9)	G (4–8)
I (Ø–9)	H (Ø–8)

(3) From the value O G it follows that G can be set next to J in E J. Thus:

. . . V H O F . . .
. . . E J G . . .

(4) But G N already is known to belong to the same set of displacement-interval values as E J. Therefore, it is now possible to combine E J, J G, and G N into one sequence, E J G N, yielding:

. . . V H O F . . .
. . . E J G N . . .

g. (1) Refer now to Figure 33.

V (Ø–22)	E (Ø–5)
? (1–22)	G (1–5)
? (2–22)	K (2–5)
? (3–22)	X (3–5)
? (5–22)	D (5–5)
? (6–22)	J (6–5)

(2) The only values which can be inserted are:

O (1–22)	G (1–5)
H (6–22)	J (6–5)

(3) This means that $V_p=O_e$ in alphabet 1 and that $V_p=H_e$ in alphabet 6. There is one O_e in the frequency distribution for alphabet 1, and no H_e in that for alphabet 6. The frequency distribution is, therefore, corroborative insofar as these values are concerned.

(*h*) (1) Further, taking E J G N and V H O F, superimpose them thus:

. . . E J G N . . .
. . . V H O F . . .

(2) Refer now to Figure 33.

E (Ø–5)	H (Ø–8)
G (1–5)	? (1–8)

(3) From the diagram of superimposition the value G (1–5) F (1–8) can be inserted, which gives $H_p=F_c$ in alphabet 1.

i. (1) Again, V H O F and E J G N are juxtaposed:

. . . V H O F . . .
. . . E J G N . . .

(2) Refer to Figure 33 and find the following:

H (Ø–8) G (4–8)
A (Ø–1) E (4–1)

This means that it is possible to add A, thus:

. . . A V H O F . . .
. . . E J G N . . .

(3) In the set there are also:

E (Ø–5) G (1–5)
G (Ø–7) Z (1–7)

Then in the superimposition

. . . E J G N . . .
. . . E J G N . . .

It is possible to add Z under G, making the sequence E J G N Z.

(4) Then taking

. . . A V H O F . . .
. . . E J G N Z . . .

and referring to Figure 33:

H (Ø–8) N (Ø–14)
O (6–8) ? (6–14)

It will be seen that O=Z from superimposition, and hence in alphabet 6 $N_p=Z_c$, an important new value, but occurring only once in the cryptogram. Has an error been made? The work so far seems too corroborative in interlocking details to think so.

j. (1) The possibilities of the superimposition and sliding of the AVHOF and the EJGNZ sequences have by no means been exhausted as yet, but a little different trail this time may be advisable.

E (Ø–5) T (Ø–20)
G (1–5) K (1–20)
K (3–5) U (3–20)

(2) Then:

. . . E J G N Z . . .
. . . T . K . . .

(3) Now refer to the following:

E (Ø–5) K (2–5)
N (Ø–14) S (2–14)

whereupon the value S can be inserted:

. . . E J G N Z . . .
. . . T . K . . S . . .

k. (1) Consider all the values based upon the displacement interval corresponding to JG:

J (6–5)	G (1–5)→	J (9– 8)	G (4– 8)				
N (6–7)	Z (1–7)	H (9– 9)	O (4– 9)				
		S (9–20)	P (4–20)→	S (2–14)	P (5–14)		
				Z (2– 8)	C (5– 8)		
				K (2– 5)	D (5– 5)		

(2) Since J and G are sequent in the E J G N Z sequence, it can be said that all the letters of the foregoing pairs are also sequent. Hence Z C, S P, and K D are available as new data. These give E J G N Z C and T . K D . S P.

(3) Now consider:

T (Ø–20)	P (4–20)
A (Ø– 1)	E (4– 1)
H (Ø– 8)	G (4– 8)
I (Ø– 9)	O (4– 9)

Now in the T . K D . S P sequence the interval between T and P is T (1 2 3 4 5 6) . . . . . P. Hence the interval between A and E is 6 also. It follows therefore that the sequences A V H O F and E J G N Z C should be united, thus:

(1 2 3 4 5 6)
. . . A V H O F . E J G N Z C . . .

(4) Corroboration is found in the interval between H and G, which is also six. The letter I can be placed into position, from the relation I (Ø–9) O (4–9), thus:

(1 2 3 4 5 6)
. . . I . . A V H O F . E J G N Z C . . .

l. (1) From Figure 33:

H (Ø– 8)	Z (2– 8)
E (Ø– 5)	K (2– 5)
N (Ø–14)	S (2–14)
U (Ø–21)	F (2–21)

(2) Since in the I . . A V H O F . E J G N Z C sequence the letters H and Z are separated by 8 intervals one can write:

(1 2 3 4 5 6 7 8)
. . . H . . . . Z . . .
. . . E . . . . K . . .
. . . N . . . . S . . .
. . . U . . . . F . . .

(3) Hence one can make the sequence

```
                                        1 2 3 4 5 6 7 8
    . . . I . . A V H O F . E J G N Z C . . K . . .
Then . . . I . . A V H O F . E J G N Z C T . K D . S P . . .
and  . . U I . . A V H O F . E J G N Z C T . K D . S P . . .
       1 2 3 4 5 6 7 8                   1 2 3 4 5 6 7 8
```

m. (1) Subsequent derivations can be indicated very briefly as follows:

```
                                  E (Ø–5)   C (Ø–3)
                                  D (5–5)   R (5–3)
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
From U I . . A V H O F . E J G N Z C T . K D . S P . . .
one can write       . . . E . . . . C . . .
                          1 2 3 4 5
and                                   . . . D . . . . R .
                                                1 2 3 4 5
making the sequence
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
      U I . . A V H O F . E J G N Z C T . K D . S P . R .
```

(2) Another derivation:

```
                                  U (3–20)  T (Ø–20)
                                  X (3– 5)  E (Ø– 5)
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
From U I . . A V H O F . E J G N Z C T . K D . S P . R .
one can write
      U I . . . . . . . . . . . . . T . . .
and   . . . . . . . . . . E . . . . . . . . . X . . . . .
making the sequence
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
      U I . . A V H O F . E J G N Z C T . K D X S P . R .
```

(3) Another derivation:

```
                                  E (Ø–5)   G (1–5)
                                  B (Ø–2)   W (1–2)

From                        . . . E J G . . .
one can write               . . . E . G . . .
and then                    . . . B . W . . .
```

There is only one place where B . W can fit, *viz*, at the end:

```
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
      U I . . A V H O F . E J G N Z C T . K D X S P B R W
```

n. Only four letters remain to be placed into the sequence, *viz*, L, M, Q, and Y. Their positions are easily found by application of the primary component to the message. The complete sequence is as follows:

```
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
      U I M Y A V H O F L E J G N Z C T Q K D X S P B R W
```

Having the primary component fully constructed, decipherment of the cryptogram can be completed with speed and precision. The text is as follows:

W F U P C F O C J Y	R C V O P N B L C W	B K D Z F M T G Q J
B U T T H O U G H W	P O S I N G T H E S	S E L F W I L L G O
G B Z D P F B O U O	L Q Z A A A M D C H	L F U Y D T Z V H Q
E C A N N O T A S Y	O L A R S Y S T E M	O U T B E C O M I N
G R F T Z M Q M A V	B Z Z C K Q O I K F	Z G W N K X J T R N
E T R E V I E W W I	S H A L L T U R N A	G A C O L D A N D L
K Z U G D Y F T R W	C F B S C V X C H Q	Y T X C D P M V L W
T H T H E M I N D S	N U N C H A N G I N	I F E L E S S M A S
G J X N L W Y O U X	Z T Z S D M X W C M	B G B W W O Q R G N
E Y E O U R P A S T	G F A C E I N P E R	S A N D T H E S O L
I T W E P Q Z O K Z	R K U H E Q E D G X	H H V L A Q Q V A V
W E C A N T O A N E	P E T U I T Y T O T	A R S Y S T E M W I
P R X C W L Z I C W	F K V H P J J K J Y	J Q W O O T T N V Q
X T E N T F O R E S	H E S U N E A C H W	L L C I R C L E U N
G K Q H O L O D V M	Y Q D P C J X L L L	B K X D S O Z R S N
E E O U R F U T U R	I L L T H E N H A V	S E E N G H O S T L
G O X S N Z H A S E	G H X E R O Q P S E	Y U X O P P Y O X Z
E W E C A N W I T H	E R E A C H E D T H	I K E I N S P A C E
B B J I P Q F J H D	G K B W T L F D U Z	H O Z O W M X C G Q
S C I E N T I F I C	E E N D O F I T S E	A W A I T I N G O N
Q C B Z E X Q T X Z	O C D H W M Z T U Z	J J U G J W Q R V M
C O N F I D E N C E	V O L U T I O N S E	L Y T H E R E S U R
J C Q R Q F V M L H	K L B P C J O T X E	U K W P E F X E N F
L O O K F O R W A R	T I N T H E U N C H	R E C T I O N O F A
S R Q E W M L N A E	H S P O P N M D L M	C C U G D W P E U H
D T O A T I M E W H	A N G I N G S T A R	N O T H E R C O S M
G S X E R O Z J S E	G C K W D V B L S E	Y B W E W V M D Y J
E N E A C H O F T H	E O F D E A T H T H	I C C A T A S T R O
G V Q W E J M K G H	G S U G D P O T H X	R Z X
E B O D I E S C O M	E N T H E S U N I T	P H E

FIGURE 34.

o. The primary component appears to be a random-mixed sequence; no key word is to be found, at least none reappears on experimentation with various hypotheses as to enciphering equations. Nevertheless, the random construction of the primary component did not complicate or retard the solution.

p. Some students may prefer to work exclusively with the reconstruction skeleton, rather than with sliding strips. One method is as good as the other and personal preferences will dictate which will be used by the individual student. If the reconstruction skeleton is used, the original letters should be inserted in ink, so as to differentiate them from derived letters.

36. General remarks.—*a.* It is to be stated that the sequence of steps described in the preceding paragraphs corresponds quite closely with that actually followed in solving the problem. It is also to be pointed out that this method can be used as a control in the early stages of analysis because it will allow the cryptanalyst to check assumptions for values. For example, the very first value derived in applying the principles of indirect symmetry to the problem herein described was $H_e=A_p$ in alphabet 1. As a matter of fact the writer had been inclined toward this value, from a study of the frequency and combinations which H_e showed; when the indirect-symmetry method actually substantiated his tentative hypothesis he immediately proceeded to substitute the value given. If he had assigned a different value to H_e, or if he had assumed a letter other than H_e for A_p in that alphabet, the conclusion would immediately follow that either the assumed value for H_e was erroneous, or that one of the values which led to the derivation of $H_e=A_p$ by indirect symmetry was wrong. Thus, these principles aid not only in the systematic and nearly automatic derivation of new values (with only occasional, or incidental references to the actual frequencies of letters), but they also assist very materially in serving as corroborative checks upon the validity of the assumptions already made.

b. Furthermore, while the writer has set forth, in the reconstruction skeleton in Figure 33, a set of 30 values apparently obtained before he began to reconstruct the primary component, this was done for purposes of clarity and brevity in exposition of the principles herein described. As a matter of fact, what he did was to watch very carefully, when inserting values in the reconstruction skeleton to find the very first chance to employ the principles of indirect symmetry; and just as soon as a value could be derived, he substituted the value in the cryptographic text. This is good procedure for two reasons. Not only will it disclose impossible combinations but also it gives opportunity for making further assumptions for values by the addition of the derived values to those previously assumed. Thus, the processes of reconstructing the primary component and finding additional data for the reconstruction proceed simultaneously in an ever-widening circle.

c. It is worth noting that the careful analysis of only 30 cipher equivalents in the reconstruction skeleton shown in Figure 33 results in the derivation of the entire table of secondary alphabets, 676 values in all. And while the elucidation of the method seems long and tedious, in its actual application the results are speedy, accurate, and gratifying in their corroborative effect upon the mental activity of the cryptanalyst.

d. (1) The problem here used as an illustrative case is by no means one that most favorably presents the application and the value of the method, for it has been applied in other cases with much speedier success. For example, suppose that in a cryptogram of 6 alphabets the equivalents of only THE in all 6 alphabets are fairly certain. As in the previous case, it is supposed that the secondary alphabets are obtained by sliding a mixed alphabet against itself. Suppose the secondary alphabets to be as follows:

Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1					B			Q												E						
2					C			L												X						
3					I			V												C						
4					N			P												B						
5					X			O												P						
6					T			Z												V						

FIGURE 35.

(2) Consider the following chain of derivatives arranged diagrammatically:

H (Ø– 8) O (5– 8)
T (Ø–20) P (5–20)
E (Ø– 5) X (5– 5)→E (1–20) X (2–20)
Q (1– 8) L (2– 8)
B (1– 5) C (2– 5)→B (4–20) C (3–20)
N (4– 5) I (3– 5)
P (4– 8) V (3– 8)→

→P (5–20) V (6–20)
O (5– 8) Z (6– 8)
X (5– 5) T (6– 5)→X (2–20) T (Ø–20)
L (2– 8) H (Ø– 8)
C (2– 5) E (Ø– 5)→C (3–20) E (1–20)
V (3– 8) Q (1– 8)
I (3– 5) B (1– 5)

FIGURE 36.

(3) These pairs manifestly all belong to the same displacement interval, and therefore unions can be made immediately. The complete list is as follows:

E X, Q L, N I, L H, H O, B C, O Z, C E, T P, P V, X T, V Q, I B

(4) Joining pairs by their common letters, the following sequence is obtained:

. . . N I B C E X T P V Q L H O Z . . .

e. With this as a nucleus the cryptogram can be solved speedily and accurately. When it is realized that the cryptanalyst can assume THE's rather readily in some cases, the value of this principle becomes apparent. When it is further realized that if a cryptogram has sufficient text to enable the THE's to be found easily, it is usually also not at all difficult to make correct assumptions of values for two or three other high-frequency letters, it is clear that the principles of indirect symmetry of position may often be used with gratifyingly quick success to reconstruct the complete primary component.

f. When the probable-word method is combined with the principles of indirect symmetry the solution of a difficult case is often accomplished with astonishing ease and rapidity.

152018—38——6

SECTION IX

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, III

	Paragraph
Solution of messages enciphered by known primary components	37
Solution of repeating-key ciphers in which the identical mixed components proceed in opposite directions	38
Solution of repeating-key ciphers in which the primary components are different mixed sequences	39
Solution of subsequent messages after the primary components have been recovered	40

37. Solution of subsequent messages enciphered by the same primary components.—*a.* In the discussion of the methods of solving repeating-key ciphers using secondary alphabets derived from the sliding of a mixed component against the normal component (Section V), it was shown how subsequent messages enciphered by the same pair of primary components but with different keys could be solved by application of principles involving the completion of the plain-component sequence (paragraphs 23, 24). The present paragraph deals with the application of these same principles to the case where the primary components are identical mixed sequences.

b. Suppose that the following primary component has been reconstructed from the analysis of a lengthy cryptogram:

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z

A new message exchanged between the same correspondents is intercepted and is suspected of having been enciphered by the same primary components but with a different key. The message is as follows:

N F W W P N O M K I W P I D S C A A E T Q V Z S E

Y O J S C A A A F G R V N H D W D S C A E G N F P

F O E M T H X L J W P N O M K I Q D B J I V N H L

T F N C S B G C R P

c. Factoring discloses that the period is 7 letters. The text is transcribed accordingly, and is as follows:

N F W W P N O
M K I W P I D
S C A A E T Q
V Z S E Y O J
S C A A A F G
R V N H D W D
S C A E G N F
P F O E M T H
X L J W P N O
M K I Q D B J
I V N H L T F
N C S B G C R
P

FIGURE 37.

d. The letters belonging to the same alphabet are then employed as the initial letters of completion sequences, in the manner shown in paragraph 23*e*, using the already reconstructed primary component. The completion diagrams for the first five letters of the first three alphabets are as follows:

ALPHABET 1					ALPHABET 2					ALPHABET 3				
N	M	S	V	S	F	K	C	Z	C	W	I	A	S	A
A	P	T	W	T	G	M	D	Q	D	X	O	B	T	B
B	R	I	X	I	H	P	F	U	F	Z	N	L	I	L
L	V	O	Z	O	J	R	G	E	G	Q	A	Y	O	Y
Y	W	N	Q	N	K	V	H	S	H	U	B	C	N	C
C	X	A	U	A	M	W	J	T	J	E	L	D	A	D
D	Z	B	E	B	P	X	K	I	K	S	Y	F	B	F
F	Q	L	S	L	R	Z	M	O	M	T	C	G	L	G
G	U	Y	T	Y	V	Q	P	N	P	I	D	H	Y	H
*H	E	C	I	C	W	U	R	A	R	O	F	J	C	J
J	S	D	O	D	X	E	V	B	V	N	G	K	D	K
K	T	F	N	F	Z	S	W	L	W	A	H	M	F	M
M	I	G	A	G	Q	T	X	Y	X	B	J	P	G	P
P	O	H	B	H	U	I	Z	C	Z	L	K	R	H	R
R	N	J	L	J	E	O	Q	D	Q	Y	M	V	S	V
V	A	K	Y	K	S	N	U	F	U	C	P	W	K	W
W	B	M	C	M	T	A	E	G	E	D	R	X	M	X
X	L	P	D	P	I	B	S	H	S	F	V	Z	P	Z
Z	Y	R	F	R	O	L	T	J	T	G	W	Q	R	Q
Q	C	V	G	V	N	Y	I	K	I	H	X	U	V	U
U	D	W	H	W	*A	C	O	M	O	J	Z	E	W	E
E	F	X	J	X	B	D	N	P	N	K	Q	S	X	X
S	G	Z	K	Z	L	F	A	R	A	M	U	T	Z	T
T	H	Q	M	Q	Y	G	B	V	B	P	E	I	Q	I
I	J	U	P	U	C	H	L	W	L	R	S	O	U	O
O	K	E	R	E	D	J	Y	X	Y	*V	T	N	E	N

FIGURE 38.

e. Examining the successive generatives to select the ones showing the best assortment of high-frequency letters, those marked in Figure 38 by asterisks are chosen. These are then assembled in columnar fashion and yield the following plain text:

1	2	3	4	5	6	7
H	A	V				
E	C	T				
C	O	N				
I	M	E				
C	O	N				

FIGURE 39.

f. The corresponding key-letters are sought, using enciphering equations $\Theta_{k/c}=\Theta_{l/p}$; $\Theta_{p/p}=\Theta_{c/c}$, and are found to be JOU, which suggests the keyword JOURNEY. Testing the key-letters RNEY for alphabets 4, 5, 6, and 7, the following results are obtained:

1	2	3	4	5	6	7
J	O	U	R	N	E	Y
N	F	W	W	P	N	O
H	A	V	E	D	I	R
S	C	A	A	E	T	Q
E	C	T	E	D	S	E

FIGURE 40.

The message may now be completed with ease. It is as follows:

J	O	U	R	N	E	Y
H	A	V	E	D	I	R
N	F	W	W	P	N	O
E	C	T	E	D	S	E
M	K	I	W	P	I	D
C	O	N	D	R	E	G
S	C	A	A	E	T	Q
I	M	E	N	T	T	O
V	Z	S	E	Y	O	J
C	O	N	D	U	C	T
S	C	A	A	A	F	G
T	H	O	R	O	R	E
R	V	N	H	D	W	D
C	O	N	N	A	I	S
S	C	A	E	G	N	F

J	O	U	R	N	E	Y
S	A	I	N	C	E	I
P	F	O	E	M	T	H
N	T	H	E	D	I	R
X	L	J	W	P	N	O
E	C	T	I	O	N	O
M	K	I	Q	D	B	J
F	H	O	R	S	E	S
I	V	N	H	L	T	F
H	O	E	F	A	L	L
N	C	S	B	G	C	R
S						
P						

FIGURE 41.

38. Solution of repeating-key ciphers in which the identical mixed components proceed in opposite directions.—The secondary alphabets in this case (paragraph 6, Case B (3) (a) (II) are reciprocal. The steps in solution are essentially the same as in the preceding case (paragraph 28); the principles of indirect symmetry of position can also be applied with the necessary modifications introduced by virtue of the reciprocity existing within the respective secondary alphabets (paragraph 31*p*).

39. Solution of repeating-key ciphers in which the primary components are different mixed sequences.—This is Case B (3) (b) of paragraph 6. The steps in solution are essentially the same as in paragraphs 28 and 31, except that in applying the principles of indirect symmetry of position it is necessary to take cognizance of the fact that the primary components are different mixed sequences (paragraph 31*q*).

40. Solution of subsequent messages after the primary components have been recovered.—*a*. In the case in which the primary components are identical mixed sequences proceeding in opposite directions, as well as in that in which the primary components are different mixed

sequences, the solution of subsequent messages[1] is a relatively easy matter. In both cases, however, the student must remember that before the method illustrated in paragraph 37 can be applied it is necessary to convert the cipher letters into their plain-component equivalents before completing the plain-component sequence. From there on, the process of selecting and assembling the proper generatrices is the same as usual.

b. Perhaps an example may be advisable. Suppose the enemy has been found to be using primary components based upon the keyword QUESTIONABLY, the plain component running from left to right, the cipher component in the reverse direction. The following new message has arrived from the intercept station:

M V X O X B Z I Y Z N L W Z H O X I E O O O E P Z
F X S R X E J B S H B O N A U R A P Z I N R A M V
X O X A I J Y X W F K N D O W J E R C U R A L V B
Z A Q U W J W X Y I D G R K D Q B D R M Q E C Y V
Q W

1	2	3	4	5	6
M	V	X	O	X	B
Z	I	Y	Z	N	L
W	Z	H	O	X	I
E	O	O	O	E	P
Z	F	X	S	R	X
E	J	B	S	H	B
O	N	A	U	R	A
P	Z	I	N	R	A
M	V	X	O	X	A
I	J	Y	X	W	F
K	N	D	O	W	J
E	R	C	U	R	A
L	V	B	Z	A	Q
U	W	J	W	X	Y
I	D	G	R	K	D
Q	B	D	R	M	Q
E	C	Y	V	Q	W

FIGURE 42.

c. Factoring discloses that the period is 6 and the message is accordingly transcribed into 6 columns, Fig. 42. The letters of these columns are then converted into their plain component equivalents by juxtaposing the two primary components at any point of coincidence, for example $Q_p = Z_c$. The converted letters are shown in Fig. 43. The letters of the individual columns are then used as the initial letters of completion sequences, using the QUESTIONABLY primary sequence. The final step is the selection and assembling of the selected generatrices. The results for the first ten letters of the first three columns are shown below:

1	2	3	4	5	6
O	S	U	M	U	H
Q	P	F	Q	K	G
E	Q	B	M	U	P
W	M	M	M	W	I
Q	Y	U	V	T	U
W	A	H	V	B	H
M	K	J	X	T	J
I	Q	P	K	T	J
O	S	U	M	U	J
P	A	F	U	E	Y
N	K	C	M	E	A
W	T	D	X	T	J
G	S	H	Q	J	Z
X	E	A	E	U	F
P	C	L	T	N	C
Z	H	C	T	O	Z
W	D	F	S	Z	E

FIGURE 43.

[1] That is, messages intercepted after the primary components have been reconstructed and enciphered by keys different from those used in the messages upon which the reconstruction of the primary components was accomplished.

Column 1									
O	Q	E	W	Q	W	M	I	O	P
N	U	S	X	U	X	P	O	N	R
A	E	T	Z	E	Z	R	N	A	V
B	S	I	Q	S	Q	V	A	B	W
L	T	O	U	T	U	W	B	L	X
Y	I	N	E	I	E	X	L	Y	Z
C	O	A	S	O	S	Z	Y	C	Q
D	N	B	T	N	T	Q	C	D	U
*F	A	L	I	A	I	U	D	F	E
G	B	Y	O	B	O	E	F	G	S
H	L	C	N	L	N	S	G	H	T
J	Y	D	A	Y	A	T	H	J	I
K	C	F	B	C	B	I	J	K	O
M	D	G	L	D	L	O	K	M	N
P	F	H	Y	F	Y	N	M	P	A
R	G	J	C	G	C	A	P	R	B
V	H	K	D	H	D	B	R	V	L
W	J	M	F	J	F	L	V	W	Y
X	K	P	G	K	G	Y	W	X	C
Z	M	R	H	M	H	C	X	Z	D
Q	P	V	J	P	J	D	Z	Q	F
U	R	W	K	R	K	F	Q	U	G
E	V	X	M	V	M	G	U	E	H
S	W	Z	P	W	P	H	E	S	J
T	X	Q	R	S	R	J	S	T	K
I	Z	U	V	Z	V	K	T	I	M

Column 2									
S	P	Q	M	Y	A	K	Q	S	A
T	R	U	P	C	B	M	U	T	B
*I	V	E	R	D	L	P	E	I	L
O	W	S	V	F	Y	R	S	O	Y
N	X	T	W	G	C	V	T	N	C
A	Z	I	X	H	D	W	I	A	D
B	Q	O	Z	J	F	X	O	B	F
L	U	N	Q	K	G	Z	N	L	G
Y	E	A	U	M	H	Q	A	Y	H
C	S	B	E	P	J	U	B	C	J
D	T	L	S	R	K	E	L	D	K
F	I	Y	T	V	M	S	Y	F	M
G	O	C	I	W	P	T	C	G	P
H	N	D	O	X	R	I	D	H	R
J	A	F	N	Z	V	O	F	J	V
K	B	G	A	Q	W	N	G	K	W
M	L	H	B	U	X	A	H	M	X
P	Y	J	L	E	Z	B	J	P	Z
R	C	K	Y	S	Q	L	K	R	Q
V	D	M	C	T	U	Y	M	V	U
W	F	P	D	I	E	C	P	W	E
X	G	R	F	O	S	D	R	X	S
Z	H	V	G	N	T	F	V	Z	T
Q	J	W	H	A	I	G	W	O	I
U	K	X	J	B	O	H	X	U	O
E	M	Z	K	L	N	J	Z	E	N

Column 3									
U	F	B	M	U	H	J	P	U	F
E	G	L	P	E	J	K	R	E	G
S	H	Y	R	S	K	M	V	S	H
T	J	C	V	T	M	P	W	T	J
I	K	D	W	I	P	R	X	I	K
O	M	F	X	O	R	V	Z	O	M
N	P	G	Z	N	V	W	Q	N	P
A	R	H	Q	A	W	X	U	A	R
B	V	J	U	B	X	Z	E	B	V
L	W	K	E	L	Z	Q	S	L	W
Y	X	M	S	Y	Q	U	T	Y	X
C	Z	P	T	C	U	E	I	C	Z
D	Q	R	I	D	E	S	O	D	Q
F	U	V	O	F	S	T	N	F	U
G	E	W	N	G	T	I	A	G	E
H	S	X	A	H	I	O	B	H	S
J	T	Z	B	J	O	N	L	J	T
K	I	Q	L	K	N	A	Y	K	I
M	O	U	Y	M	A	B	C	M	O
P	N	E	C	P	B	L	D	P	N
*R	A	S	D	R	L	Y	F	R	A
V	B	T	F	V	Y	C	G	V	B
W	L	I	G	W	C	D	H	W	L
X	Y	O	H	X	D	F	J	X	Y
Z	C	N	J	Z	F	G	K	Z	C
Q	D	A	K	Q	G	H	N	Q	D

FIGURE 44.

Columnar assembling of selected generatrices gives what is shown in Fig. 45.

1	2	3	4	5	6
F	I	R	.	.	.
A	V	A	.	.	.
L	E	S	.	.	.
I	R	D	.	.	.
A	D	R	.	.	.
I	L	L	.	.	.
U	P	Y	.	.	.
D	E	F	.	.	.
F	I	R	.	.	.
E	L	A	.	.	.

FIGURE 45.

d. The key letters are sought, and found to be NUM, which suggests NUMBER. The entire message may now be read with ease. It is as follows:

N	U	M	B	E	R
F	I	R	S	T	C
M	V	X	O	X	B
A	V	A	L	R	Y
Z	I	Y	Z	N	L
L	E	S	S	T	H
W	Z	H	O	X	I
I	R	D	S	Q	U
E	O	O	O	E	P
A	D	R	O	N	W
Z	F	X	S	R	X
I	L	L	O	C	C
E	J	B	S	H	B
U	P	Y	A	N	D
O	N	A	U	R	A
D	E	F	E	N	D
P	Z	I	N	R	A
F	I	R	S	T	D
M	V	X	O	X	A

N	U	M	B	E	R
E	L	A	Y	I	N
I	J	Y	X	W	F
G	P	O	S	I	T
K	N	D	O	W	J
I	O	N	A	N	D
E	R	C	U	R	A
W	I	L	L	P	R
L	V	B	Z	A	Q
O	T	E	C	T	L
U	W	J	W	X	Y
E	F	T	F	L	A
I	D	G	R	K	D
N	K	O	F	B	R
Q	B	D	R	M	Q
I	G	A	D	E	X
E	C	Y	V	Q	W

FIGURE 46.

e. If the primary components are different mixed sequences, the procedure is identical with that just indicated. The important point to note is that one must not fail to convert the letters into their plain-component equivalents before the completion-sequence method is applied.

SECTION X

REPEATING-KEY SYSTEMS WITH MIXED CIPHER ALPHABETS, IV

	Paragraph
General remarks	41
Deriving the secondary alphabets, the primary components, and the key, given a cryptogram with its plain text	42
Deriving the secondary alphabets, the primary components, and the keywords for messages, given two or more cryptograms in different keys and suspected to contain identical plain text	43
The case of repeating-key systems	44
The case of identical messages enciphered by keywords of different lengths	45
Concluding remarks	46

41. General remarks.—The preceding three sections have been devoted to an elucidation of the general principles and procedure in the solution of typical cases of repeating-key ciphers. This section will be devoted to a consideration of the variations in cryptanalytic procedure arising from special circumstances. It may be well to add that by the designation "special circumstances" it is not meant to imply that the latter are necessarily *unusual* circumstances. *The student should always be on the alert to seize upon any opportunities that may appear in which he may apply the methods to be described.* In practical work such opportunities are by no means rare and are seldom overlooked by competent cryptanalysts.

42. Deriving the secondary alphabets, the primary components, and the key, given a cryptogram with its plain text.—*a.* It may happen that a cryptogram and its equivalent plain text are at hand, as the result of capture, pilferage, compromise, etc. This, as a general rule, affords a very easy attack upon the whole system.

b. Taking first the case where the plain component is the normal alphabet, the cipher component a mixed sequence, the first thing to do is to write out the cipher text with its letter-for-letter decipherment. From this, by a slight modification of the principles of "factoring", one discovers the length of the key. It is obvious that when a word of three or four letters is enciphered by the same cipher text, the interval between the two occurrences is almost certainly a multiple of the length of the key. By noting a few recurrences of plain text and cipher letters, one can quickly determine the length of the key (assuming of course that the message is long enough to afford sufficient data). Having determined the length of the key, the message is rewritten according to its periods, with the plain text likewise in periods under the cipher letters. From this arrangement one can now reconstruct complete or partial secondary alphabets. If the secondary alphabets are complete, they will show direct symmetry of position; if they are but fragmentary in several alphabets, then the primary component can be reconstructed by the application of the principles of direct symmetry of position.

c. If the plain component is a mixed sequence, and the cipher component the normal (direct or reversed sequence), the secondary alphabets will show no direct symmetry unless they are arranged in the form of deciphering alphabets (that is, $A_c \ldots Z_c$ above the zero line, with their equivalents below). The student should be on the lookout for such cases.

d. (1) If the plain and cipher primary components are identical mixed sequences proceeding in the same direction, the secondary alphabets will show indirect symmetry of position, and they can be used for the speedy reconstruction of the primary components (Paragraph 31*a* to *o*).

(2) If the plain and the cipher primary components are identical mixed sequences proceeding in opposite directions, the secondary alphabets will be completely reciprocal secondary alphabets and the primary component may be reconstructed by applying the principles outlined in paragraph 31*p*.

(3) If the plain and the cipher primary components are different mixed sequences, the secondary alphabets will show indirect symmetry of position and the primary components may be reconstructed by applying the principles outlined in paragraph 31*q*.

e. In all the foregoing cases, after the primary components have been reconstructed, the keys can be readily recovered.

43. Deriving the secondary alphabets, the primary components, and the keywords for messages, given two or more cryptograms in different keys and suspected to contain identical plain text.—*a*. The simplest case of this kind is that involving two monoalphabetic substitution ciphers with mixed alphabets derived from the same pair of sliding components. An understanding of this case is necessary to that of the case involving repeating-key ciphers.

b. (1) A message is transmitted from station A to station B. B then sends A some operating signals which indicate that B cannot decipher the message, and soon thereafter A sends a second message, identical in length with the first. This leads to the suspicion that the plain text of both messages is the same. The intercepted messages are superimposed. Thus:

1. NXGRV MPUOF ZQVCP VWERX QDZVX WXZQE TBDSP VVXJK RFZWH ZUWLU IYVZQ FXOAR
2. EMLHJ FGVUB PRJNG JKWHM RAPJM KMPRW ZTAXG JJMCD HBPKY PVKIV QOJPR BMUSH

(2) Initiating a chain of cipher-text equivalents from message 1 to message 2, the following complete sequence is obtained:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
N	E	W	K	D	A	S	X	M	F	B	T	Z	P	G	L	I	Q	R	H	Y	O	U	V	J	C

(3) Experimentation along already-indicated lines soon discloses the fact that the foregoing component is an equivalent primary component of the original primary based upon the keyword QUESTIONABLY, decimated on the 21st interval. Let the student decipher the cryptogram.

(4) The foregoing example is somewhat artificial in that the plain text was consciously selected with a view to making it contain every letter of the alphabet. The purpose in doing this was to permit the construction of a complete chain of equivalents from only two short messages, in order to give a simple illustration of the principles involved. If the plain-text message does not contain every letter of the alphabet, then only partial chains of equivalents can be constructed. These may be united, if circumstances will permit, by recourse to the various principles elucidated in paragraph 31.

(5) The student should carefully study the foregoing example in order to obtain a thorough comprehension of the *reason* why it was possible to reconstruct the primary component from the two cipher messages without having any plain text to begin with at all. Since the plain text of both messages is the same, the relative displacement of the primary components in the case of message 1 differs from the relative displacement of the same primary components in the case of message 2 by a *fixed* interval. Therefore, the distance between N and E (the first letters of the two messages), on the primary component, regardless of what plain-text letter these two cipher letters represent, is the same as the distance between E and W (the 18th letters), W and K (the 17th letters), and so on. Thus, this fixed interval permits of establishing a complete chain of letters separated by constant intervals and this chain becomes an equivalent primary component.

44. The case of repeating-key systems.—*a.* With the foregoing basic principles in mind the student is ready to note the procedure in the case of two repeating-key ciphers having identical plain texts. First, the case in which both messages have keywords of identical length but different compositions will be studied.

b. (1) Given the following two cryptograms suspected to contain the same plain text:

MESSAGE 1

```
Y H Y E X   U B U K A   P V L L T   A B U V V   D Y S A B
P C Q T U   N G K F A   Z E F I Z   B D J E Z   A L V I D
T R O Q S   U H A F K
```

MESSAGE 2

```
C G S L Z   Q U B M N   C T Y B V   H L Q F T   F L R H L
M T A I Q   Z W M D Q   N S D W N   L C B L Q   N E T O C
V S N Z R   B J N O Q
```

(2) The first step is to try to determine the length of the period. The usual method of factoring cannot be employed because there are no long repetitions and not enough repetitions even of digraphs to give any convincing indications. However, a subterfuge will be employed, based upon the theory of factoring.

c. (1) Let the two messages be superimposed.

```
    1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
1.  Y  H  Y  E  X  U  B  U  K  A  P  V  L  L  T  A  B  U  V  V
2.  C  G  S  L  Z  Q  U  B  M  N  C  T  Y  B  V  H  L  Q  F  T

   21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
1.  D  Y  S  A  B  P  C  Q  T  U  N  G  K  F  A  Z  E  F  I  Z
2.  F  L  R  H  L  M  T  A  I  Q  Z  W  M  D  Q  N  S  D  W  N

   41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
1.  B  D  J  E  Z  A  L  V  I  D  T  R  O  Q  S  U  H  A  F  K
2.  L  C  B  L  Q  N  E  T  O  C  V  S  N  Z  R  B  J  N  O  Q
```

```
                                                          4     44
                                                          E     E
```

(2) Now let a search be made of cases of identical superimposition. For example, L and L

```
      6  18        30
      U  U         U
```

are separated by 40 letters, Q, Q, and Q are separated by 12 letters. Let these intervals between identical superimpositions be factored, just as though they were ordinary repetitions. That factor which is the most frequent should correspond with the length of the period for the following reason. If the period is the same and the plain text is the same in both messages, then the condition of identity of superimposition can only be the result of identity of encipherments by identical cipher alphabets. This is only another way of saying that the same relative position in the keying cycle has been reached in both cases of identity. Therefore, the distance between identical superimpositions must be either equal to or else a multiple of the length of the period. Hence, factoring the intervals must yield the length of the period. The complete list of intervals

and factors applicable to cases of identical superimposed pairs is as follows (factors above 12 are omitted):

Repetition	Interval	Factors	Repetition	Interval	Factors
1st EL to 2d EL	40	2, 4, 5, 8, 10.	1st TV to 2d TV	36	2, 3, 4, 6, 9, 12.
1st UQ to 2d UQ	12	2, 3, 4, 6, 12.	1st AH to 2d AH	8	2, 4, 8.
2d UQ to 3d UQ	12	2, 3, 4, 6, 12.	1st BL to 2d BL	8	2, 4, 8.
1st UB to 2d UB	48	2, 3, 4, 6, 8, 12.	2d BL to 3d BL	16	2, 4, 8.
1st KM to 2d KM	24	2, 3, 4, 6, 8, 12.	1st SR to 2d SR	32	2, 4, 8.
1st AN to 2d AN	36	2, 3, 4, 6, 9, 12.	1st FD to 2d FD	4	2, 4.
2d AN to 3d AN	12	2, 3, 4, 6, 12.	1st ZN to 2d ZN	4	2, 4.
1st VT to 2d VT	8	2, 4, 8.	1st DC to 2d DC	8	2, 4, 8.
2d VT to 3d VT	28	2, 4, 7.			

(3) The factor 4 is the only one common to every one of these intervals and it may be taken as beyond question that the length of the period is 4.

d. Let the messages now be superimposed according to their periods:

	1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4
1.	Y	H	Y	E		X	U	B	U		K	A	P	V		L	L	T	A		B	U	V	V		D	Y	S	A		B	P	C	Q
2.	C	G	S	L		Z	Q	U	B		M	N	C	T		Y	B	V	H		L	Q	F	T		F	L	R	H		L	M	T	A
1.	T	U	N	G		K	F	A	Z		E	F	I	Z		B	D	J	E		Z	A	L	V		I	D	T	R		O	Q	S	U
2.	I	Q	Z	W		M	D	Q	N		S	D	W	N		L	C	B	L		Q	N	E	T		O	C	V	S		N	Z	R	B
1.	H	A	F	K																														
2.	J	N	O	Q																														

e. (1) Now distribute the superimposed letters into a reconstruction skeleton of "secondary alphabets."
Thus:

Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1		L		F	S			J	O		M	Y			N					I					Z	C	Q
2	N			C		D		G				B				M	Z				Q				L	
3	Q	U	T			O			W	B		E		Z		C			R	V		F			S	
4	H				L		W				Q						A	S			B	T				N

(2) By the usual methods, construct the primary or an equivalent primary component. Taking lines Ø and 1, the following sequences are noted:

BL, DF, ES, HJ, IO, KM, LY, ON, TI, XZ, YC, ZQ,

which, when united by means of common letters and study of other sequences, yield the complete original primary component based upon the keyword QUESTIONABLY:

Q U E S T I O N A B L Y C D F G H J K M P R V W X Z

(3) The fact that the pair of lines with which the process was commenced yield the original primary sequence is purely accidental; it might have just as well yielded an equivalent primary sequence.

f. (1) Having the primary component, the solution of the messages is now a relatively simple matter. An application of the method elucidated in paragraph 37 is made, involving the completion of the plain-component sequence for each alphabet and selecting those generatrices which contain the best assortments of high-frequency letters. Thus, using Message 1:

First Alphabet	Second Alphabet	Third Alphabet	Fourth Alphabet
Y X K L B	H U A L U	Y B P T V	E U V A V
C Z M Y L	J E B Y E	C L R I W	S E W B W
D Q P C Y	K S L C S	D Y V O X	T S X L X
F U R D C	M T Y D T	F C W N Z	I T Z Y Z
G E V F D	P I C F I	G D X A Q	O I Q C Q
H S W G F	R O D G O	H F Z B U	N O U D U
J T X H G	V N F H N	J G Q L E	*A N E F E
K I Z J H	W A G J A	K H U Y S	B A S G S
M O Q K J	X B H K B	M J E C T	L B T H T
P N U M K	Z L J M L	P K S D I	Y L I J I
R A E P M	Q Y K P Y	R M T F O	C Y O K O
V B S R P	U C M R C	V P I G N	D C N M N
W L T V R	E D P V D	W R O H A	F D A P A
X Y I W V	S F R W F	X V N J B	G F B R B
Z C O X W	T G V X G	Z W A K L	H G L V L
Q D N Z X	I H W Z H	Q X B M Y	J H Y W Y
U F A Q Z	O J X Q J	U Z L P C	K J C X C
E G B U Q	N K Z U K	E Q Y R D	M K D Z D
S H L E U	A M Q E M	S U C V F	P M F Q F
T J Y S E	B P U S P	T E D W G	R P G U G
I K C T S	*L R E T R	I S F X H	V R H E H
O M D I T	Y V S I V	O T G Z J	W V J S J
N P F O I	C W T O W	N I H Q K	X W K T K
*A R G N O	D X I N X	A O J U M	Z X M I M
B V H A N	F Z O A Z	B N K E P	Q Z P O P
L W J B A	G Q N B Q	*L A M S R	U Q R N R

Figure 48.

(2) The selected generatrices (those marked by asterisks in Fig. 48) are assembled in columnar manner:

A L L A
R R A N
G E M E
N T S F
O R R E

Figure 49.

(3) The key letters are sought and give the keyword SOUP. The plain text for the second message is now known, and by reference to the cipher text and the primary components, the keyword for this message is found to be TIME. The complete texts are as follows:

S O U P	T I M E
A L L A	A L L A
Y H Y E	C G S L
R R A N	R R A N
X U B U	Z Q U B
G E M E	G E M E
K A P V	M N C T
N T S F	N T S F
L L T A	Y B V H
O R R E	O R R E
B U V V	L Q F T
L I E F	L I E F
D Y S A	F L R H
O F Y O	O F Y O
B P C Q	L M T A
U R O R	U R O R
T U N G	I Q Z W
G A N I	G A N I
K F A Z	M D Q N
Z A T I	Z A T I
E F I Z	S D W N
O N H A	O N H A
B D J E	L C B L
V E B E	V E B E
Z A L V	Q N E T
E N S U	E N S U
I D T R	O C V S
S P E N	S P E N
O Q S U	N Z R B
D E D X	D E D X
H A F K	J N O Q

FIGURE 50.

45. **The case of identical messages enciphered by keywords of different lengths.**—*a.* In the foregoing case the keywords for the two messages, although different, were identical in length. When this is not true and the keywords are of different lengths, the procedure need be only slightly modified.

b. Given the following two cryptograms suspected of containing the same plain-text enciphered by the same primary components but with different keywords of different lengths, solve the messages.

MESSAGE No. 1

V M Y Z G E A U N T P K F A Y J I Z M B U M Y K B V F I V V
S E O A F S K X K R Y W C A C Z O R D O Z R D E F B L K F E
S M K S F A F E K V Q U R C M Y Z V O X V A B T A Y Y U O A
Y T D K F E N W N T D B Q K U L A J L Z I O U M A B O A F S
K X Q P U Y M J P W Q T D B T O S I Y S M I Y K U R O G M W
C T M Z Z V M V A J

MESSAGE No. 2

Z G A M W I O M O A C O D H A C L R L P M O Q O J E M O Q U
D H X B Y U Q M G A U V G L Q D B S P U O A B I R P W X Y M
O G G F T M R H V F G W K N I V A U P F A B R V I L A Q E M
Z D J X Y M E D D Y B O S V M P N L G X X D Y D O P X B Y U
Q M N K Y F L U Y Y G V P V R D N C Z E K J Q O R W J X R V
G D K D S X C E E C

c. The messages are long enough to show a few short repetitions which permit factoring. The latter discloses that Message 1 has a period of 4 and Message 2, a period of 6 letters. The messages are superimposed, with numbers marking the position of each letter in the corresponding period, as shown below:

```
                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. V M Y Z G E A U N T P K F A Y J I Z M B U M Y K
No. 2. Z G A M W I O M O A C O D H A C L R L P M O Q O
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. B V F I V V S E O A F S K X K R Y W C A C Z O R
No. 2. J E M O Q U D H X B Y U Q M G A U V G L Q D B S
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. D O Z R D E F B L K F E S M K S F A F E K V Q U
No. 2. P U O A B I R P W X Y M O G G F T M R H V F G W
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. R C M Y Z V O X V A B T A Y Y U O A Y T D K F E
No. 2. K N I V A U P F A B R V I L A Q E M Z D J X Y M
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. N W N T D B Q K U L A J L Z I O U M A B O A F S
No. 2. E D D Y B O S V M P N L G X X D Y D O P X B Y U
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. K X Q P U Y M J P W Q T D B T O S I Y S M I Y K
No. 2. Q M N K Y F L U Y Y G V P V R D N C Z E K J Q O
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

                1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
No. 1. U R O G M W C T M Z Z V M V A J
No. 2. R W J X R V G D K D S X C E E C
       1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4
```

d. A reconstruction skeleton of "secondary alphabets" is now made by distributing the letters in respective lines corresponding to the 12 different superimposed pairs of numbers. For example, all pairs corresponding to the superimposition of position 1 of Message 1 with position 1 of Message 2 are distributed in lines ∅ and 1 of the skeleton. Thus, the very first superimposed pair is $\begin{cases} 1 \\ V \\ Z \\ 1 \end{cases}$; the letter Z is inserted in line 1 under the letter V. The next $\begin{cases} 1 \\ 1 \end{cases}$ pair is the 13th superimposition, with $\begin{cases} F \\ D \end{cases}$; the letter D is inserted in line 1 under the letter F, and so on. The skeleton is then as follows:

Superimposed pairs. ∅	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1–1	I	J		P		D					Q	G	C	E				K	O		R	Z				
2–2	H	V	N										G		U			W				E	D	M	L	X
3–3	E					M			X		G		I	D	J		N			R					A	O
4–4							X		O	C					D	K		A	F	Y	Q				V	M
1–5				B		T	W		L				R		E				N		Y	Q			U	A
2–6	M	O			I				C				D									U	V		F	R
3–1	O		G			R							L		P		S		D						Z	
4–2	L	P			H					U	V								E	D	M			F		
1–3			Q	J							V	W	K	O	X	Y					M	A				
2–4	B								J		X	P	O							A		F	Y			D
3–5	N	R				Y									B	C	G								Q	S
4–6					M					L	O							S	U	V	W	X				

FIGURE 51.

e. There are more than sufficient data here to permit of the reconstruction of a complete equivalent primary component, for example, the following:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
I	T	K	N	P	Z	H	M	W	B	Q	E	U	L	F	C	S	J	A	X	R	G	D	V	O	Y

f. The subsequent steps in the actual decipherment of the text of either of the two messages are of considerable interest. Thus far the cryptanalyst has only the cipher component of the primary sliding components. The plain component may be identical with the cipher component and may progress in the same direction, or in the reverse direction; or, the two components may be different. If different, the plain component may be the normal sequence, direct or reversed. Tests must be made to ascertain which of these various possibilities is true.

g. (1) It will first be assumed that the primary plain component is the normal direct sequence. Applying the procedure outlined in Par. 23 to the message with the shorter key (Message No. 1, to give the most data per secondary alphabet), an attempt is made to solve the message. It is unnecessary here to go further into detail in this procedure; suffice it to indicate that the attempt is unsuccessful and it follows that the plain component is not the normal direct sequence. A normal reversed sequence is then assumed for the plain component and the proper procedure applied. Again the attempt is found useless. Next, it is assumed that the plain component is identical with the cipher component, and the procedure outlined in Par. 37 is tried. This also is unsuccessful. Another attempt, assuming the plain component runs in the reverse direction, is likewise unsuccessful. There remains one last hypothesis, viz, that the two primary components are different mixed sequences.

(2) Here is Message No. 1 transcribed in periods of four letters. Uniliteral frequency distributions for the four secondary alphabets are shown below in Fig. 52, labeled 1*a*, 2*a*, 3*a*, and 4*a*. These distributions are based upon the normal sequence A to Z. But since the reconstructed cipher component is at hand these distributions can be rearranged according to the sequence of the cipher component, as shown in distributions labeled 1*b*, 2*b*, 3*b*, and 4*b* in Fig. 52. *The latter distributions may be combined by shifting distributions 2b, 3b, and 4b to proper superimpositions with respect to 1b so as to yield a single monoalphabetic distribution for the entire message. In other words, the polyalphabetic message can be converted into monoalphabetic terms, thus very considerably simplifying the solution.*

MESSAGE No. 1

V M Y Z V A B T
G E A U A Y Y U
N T P K O A Y T
F A Y J D K F E
I Z M B N W N T
U M Y K D B Q K
B V F I U L A J
V V S E L Z I O
O A F S U M A B
K X K R O A F S
Y W C A K X Q P
C Z O R U Y M J
D O Z R P W Q T
D E F B D B T O
L K F E S I Y S
S M K S M I Y K
F A F E U R O G
K V Q U M W C T
R C M Y M Z Z V
Z V O X M V A J

1*a*. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

2*a*. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

3*a*. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

4*a*. A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

1*b*. I T K N P Z H M W B Q E U L F C S J A X R G D V O Y

2*b*. I T K N P Z H M W B Q E U L F C S J A X R G D V O Y

3*b*. I T K N P Z H M W B Q E U L F C S J A X R G D V O Y

4*b*. I T K N P Z H M W B Q E U L F C S J A X R G D V O Y

FIGURE 52.

(3) Note in Fig. 53 how the four distributions are shifted for superimposition and how the combined distribution presents the characteristics of a typical monoalphabetic distribution.

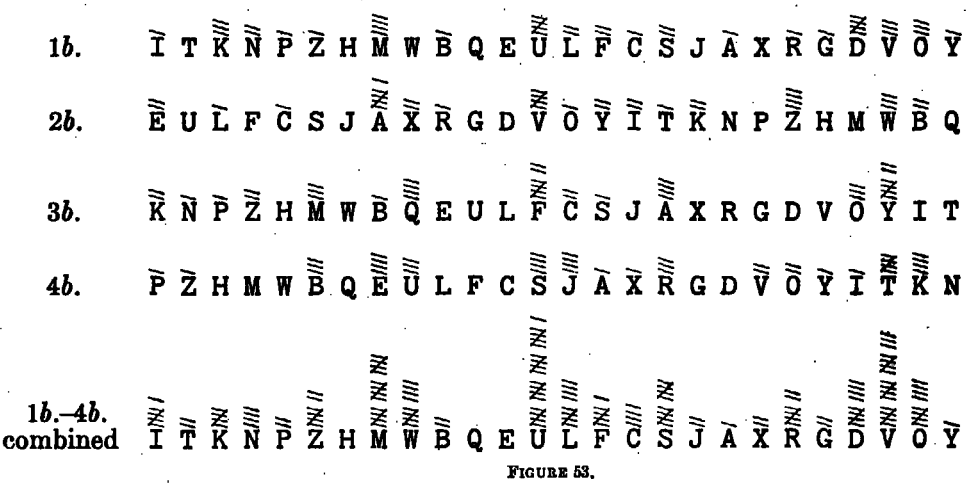


FIGURE 53.

(4) The letters belonging to alphabets 2, 3, and 4 of Fig. 52 may now be transcribed in terms of alphabet 1. That is, the two E's of alphabet 2 become I's; the L of alphabet 2 becomes a K; the C becomes a P, and so on. Likewise, the two K's of alphabet 3 become I's, the N becomes a T, and so on. The entire message is then a monoalphabet and can readily be solved. It is as follows:

V D V T G	I S W N S	K O F M V	L I R Z Z	U D V O B	U U D V U
E N E M Y	H A S C A	P T U R E	D H I L L	O N E T W	O O N E O
F M O M U	U K W I S	Y V L F C	R D S D L	N S D I U	Z L J U M
U R T R O	O P S H A	V E D U G	I N A N D	C A N H O	L D F O R
S D I U F	M U M K U	W W R P Z	G Z U D C	V M M V A	F V W O M
A N H O U	R O R P O	S S I B L	Y L O N G	E R R E Q	U E S T R
V V D J U	M N V T V	D O W O U	K S L L R	O R U D S	Z O M U U
E I N F O	R C E M E	N T S T O	P A D D I	T I O N A	L T R O O
K W W I U	F Z L P V	W V D O Y	R S C V U	M C V O U	B D J M V
P S S H O	U L D B E	S E N T V	I A G E O	R G E T O	W N F R E
L V M R N	X M U S L				
D E R I C	K R O A D				

(5) Having the plain text, the derivation of the cipher component (an equivalent) is an easy matter. It is merely necessary to base the reconstruction upon any of the secondary alphabets, since the plain text—cipher relationship is now known directly, and the primary cipher component is at hand. The primary plain component is found to be as follows:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
H	M	P	C	B	L	.	R	S	W	.	.	O	D	U	G	A	F	Q	K	I	Y	N	E	T	V

(6) The keywords for both messages can now be found, if desirable, by finding the equivalent of A_p in each of the secondary alphabets of the original polyalphabetic messages. The keyword for No. 1 is STAR; that for No. 2 is OCEANS.

152018—38——7

(7) The student may, if he wishes, try to find out whether the primary components reconstructed above are the original components or are equivalent components, by examining all the possible decimations of the two components for evidences of derivation from keywords.

h. As already stated in Par. 26*l*, there are certain statistical and mathematical tests that can be employed in the process of "matching" distributions to ascertain proper superimpositions for monoalphabeticity. In the case just considered there were sufficient data in the distributions to permit the process to be applied successfully by eye, without necessitating statistical tests.

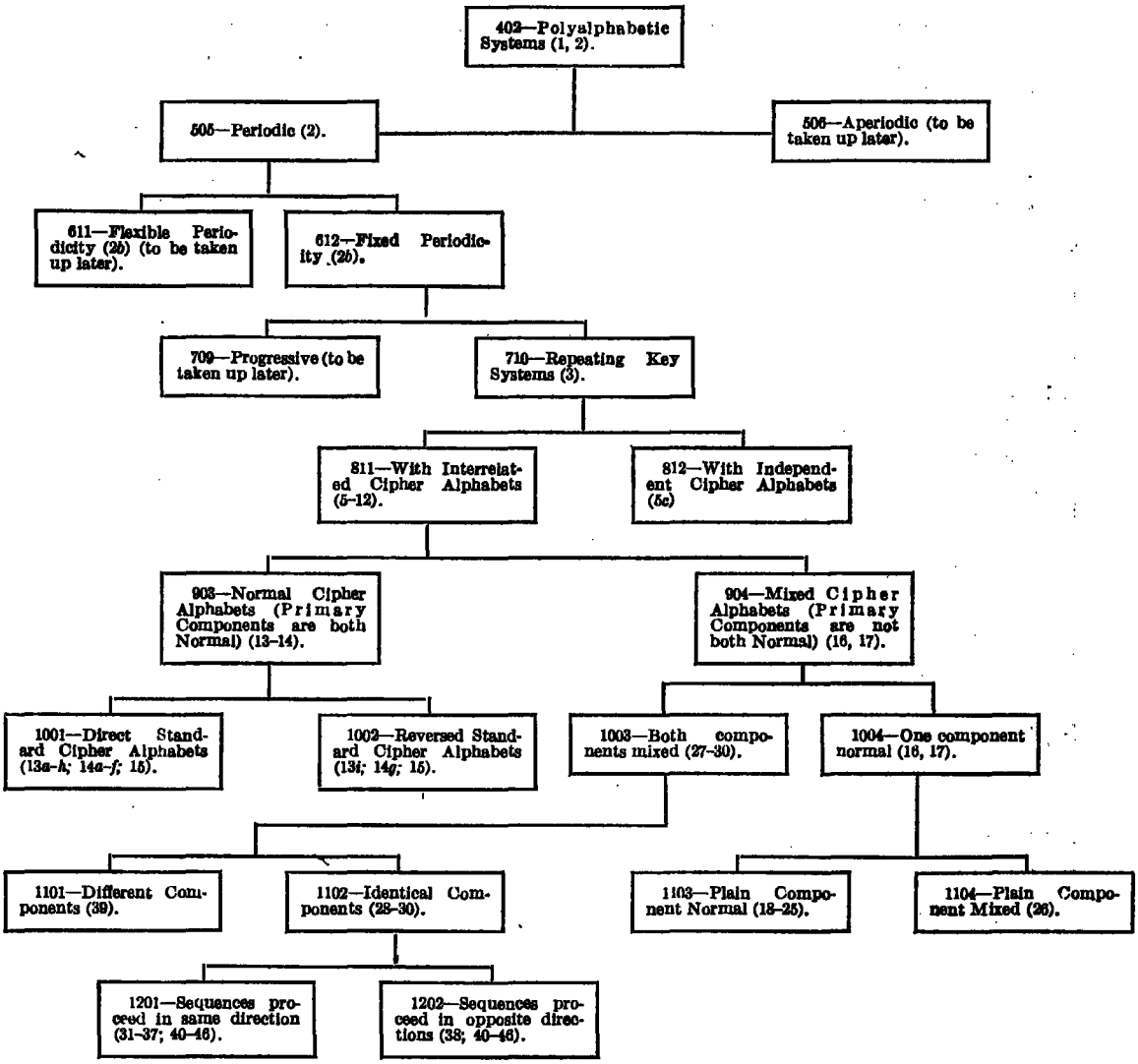
i. This case is an excellent illustration of the application of the process of *converting a polyalphabetic cipher into monoalphabetic terms.* Because it is a very valuable and important cryptanalytic "trick," the student should study it most carefully in order to gain a good understanding of the principle upon which it is based and its significance in cryptanalysis. The conversion in the case under discussion was possible because the sequence of letters forming the cipher component had been reconstructed and was known, and therefore the uniliteral distributions for the respective secondary cipher alphabets could theoretically be shifted to correct superimpositions for monoalphabeticity. It also happened that there were sufficient data in the distributions to give proper indications for their relative displacements. Therefore, the theoretical possibility in this case became an actuality. Without these two necessary conditions the superimposition and conversion cannot be accomplished. The student should always be on the lookout for situations in which this is possible.

46. Concluding remarks.—*a.* The observant student will have noted that a large part of this text is devoted to the elucidation and application of a very few basic principles. These principles are, however, extremely important and their proper usage in the hands of a skilled cryptanalyst makes them practically indispensable tools of his art. The student should therefore drill himself in the application of these tools by having someone make up problem after problem for him to practice upon, until he acquires facility in their use and feels competent to apply them in practice whenever the least opportunity presents itself. This will save him much time and effort in the solution of bona fide messages.

b. Continuing the analytical key introduced in Military Cryptanalysis Part I, the outline for the studies covered by Part II follows herewith.

Analytical Key for Military Cryptanalysis, Part II *

(Numbers in parentheses refer to Paragraph Numbers in this text)



*For explanation of the use of this chart see Par. 50 of Military Cryptanalysis, Part I.

APPENDIX 1

THE 12 TYPES OF CIPHER SQUARES

(See Paragraph 7)

TABLE I–B.[1]

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/2}=\theta_{l/1}$; $\theta_{p/1}=\theta_{c/2}$ ($\theta_{l/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V
B	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F
C	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R
D	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T
E	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S
F	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X
G	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I
H	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E
I	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z
J	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D
K	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M
L	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A
M	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U
N	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W
O	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N
P	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B
Q	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C
R	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y
S	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G
T	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H
U	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J
V	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K
W	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L
X	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O
Y	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P
Z	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q

[1] This table is labeled "Table 1–B" because it is the same as Table 1–A on page 7, except that the horizontal lines of the latter have been shifted so as to begin the successive alphabets with the successive letters of the normal sequence.

TABLE II

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/2}=\Theta_{l/1}$; $\Theta_{p/2}=\Theta_{c/1}$ ($\Theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	H	L	U	R	G	P	S	O	V	Y	B	X	D	E	I	M	K	Q	T	W	Z	C	F	J	N
B	T	A	E	N	K	Z	I	L	H	O	R	U	Q	W	X	B	F	D	J	M	P	S	V	Y	C	G
C	P	W	A	J	G	V	E	H	D	K	N	Q	M	S	T	X	B	Z	F	I	L	O	R	U	Y	C
D	G	N	R	A	X	M	V	Y	U	B	E	H	D	J	K	O	S	Q	W	Z	C	F	I	L	P	T
E	J	Q	U	D	A	P	Y	B	X	E	H	K	G	M	N	R	V	T	Z	C	F	I	L	O	S	W
F	U	B	F	O	L	A	J	M	I	P	S	V	R	X	Y	C	G	E	K	N	Q	T	W	Z	D	H
G	L	S	W	F	C	R	A	D	Z	G	J	M	I	O	P	T	X	V	B	E	H	K	N	Q	U	Y
H	I	P	T	C	Z	O	X	A	W	D	G	J	F	L	M	Q	U	S	Y	B	E	H	K	N	R	V
I	M	T	X	G	D	S	B	E	A	H	K	N	J	P	Q	U	Y	W	C	F	I	L	O	R	V	Z
J	F	M	Q	Z	W	L	U	X	T	A	D	G	C	I	J	N	R	P	V	Y	B	E	H	K	O	S
K	C	J	N	W	T	I	R	U	Q	X	A	D	Z	F	G	K	O	M	S	V	Y	B	E	H	L	P
L	Z	G	K	T	Q	F	O	R	N	U	X	A	W	C	D	H	L	J	P	S	V	Y	B	E	I	M
M	D	K	O	X	U	J	S	V	R	Y	B	E	A	G	H	L	P	N	T	W	Z	C	F	I	M	Q
N	X	E	I	R	O	D	M	P	L	S	V	Y	U	A	B	F	J	H	N	Q	T	W	Z	C	G	K
O	W	D	H	Q	N	C	L	O	K	R	U	X	T	Z	A	E	I	G	M	P	S	V	Y	B	F	J
P	S	Z	D	M	J	Y	H	K	G	N	Q	T	P	V	W	A	E	C	I	L	O	R	U	X	B	F
Q	O	V	Z	I	F	U	D	G	C	J	M	P	L	R	S	W	A	Y	E	H	K	N	Q	T	X	B
R	Q	X	B	K	H	W	F	I	E	L	O	R	N	T	U	Y	C	A	G	J	M	P	S	V	Z	D
S	K	R	V	E	B	Q	Z	C	Y	F	I	L	H	N	O	S	W	U	A	D	G	J	M	P	T	X
T	H	O	S	B	Y	N	W	Z	V	C	F	I	E	K	L	P	T	R	X	A	D	G	J	M	Q	U
U	E	L	P	Y	V	K	T	W	S	Z	C	F	B	H	I	M	Q	O	U	X	A	D	G	J	N	R
V	B	I	M	V	S	H	Q	T	P	W	Z	C	Y	E	F	J	N	L	R	U	X	A	D	G	K	O
W	Y	F	J	S	P	E	N	Q	M	T	W	Z	V	B	C	G	K	I	O	R	U	X	A	D	H	L
X	V	C	G	P	M	B	K	N	J	Q	T	W	S	Y	Z	D	H	F	L	O	R	U	X	A	E	I
Y	R	Y	C	L	I	X	G	J	F	M	P	S	O	U	V	Z	D	B	H	K	N	Q	T	W	A	E
Z	N	U	Y	H	E	T	C	F	B	I	L	O	K	Q	R	V	Z	X	D	G	J	M	P	S	W	A

TABLE III

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\theta_{k/1}=\theta_{1/2}$; $\theta_{p/1}=\theta_{c/2}$ ($\theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X
B	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O
C	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N
D	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W
E	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L
F	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A
G	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V
H	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K
I	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M
J	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U
K	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J
L	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D
M	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H	T
N	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E	H
O	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S	E
P	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G	S
Q	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I	G
R	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z	I
S	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q	Z
T	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C	Q
U	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R	C
V	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y	R
W	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P	Y
X	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B	P
Y	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F	B
Z	B	P	Y	R	C	Q	Z	I	G	S	E	H	T	D	J	U	M	K	V	A	L	W	N	O	X	F

TABLE IV

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/l}=\Theta_{1/2}$; $\Theta_{p/2}=\Theta_{o/l}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	U	B	F	O	L	A	J	M	I	P	S	V	R	X	Y	C	G	E	K	N	Q	T	W	Z	D	H
B	V	C	G	P	M	B	K	N	J	Q	T	W	S	Y	Z	D	H	F	L	O	R	U	X	A	E	I
C	W	D	H	Q	N	C	L	O	K	R	U	X	T	Z	A	E	I	G	M	P	S	V	Y	B	F	J
D	X	E	I	R	O	D	M	P	L	S	V	Y	U	A	B	F	J	H	N	Q	T	W	Z	C	G	K
E	Y	F	J	S	P	E	N	Q	M	T	W	Z	V	B	C	G	K	I	O	R	U	X	A	D	H	L
F	Z	G	K	T	Q	F	O	R	N	U	X	A	W	C	D	H	L	J	P	S	V	Y	B	E	I	M
G	A	H	L	U	R	G	P	S	O	V	Y	B	X	D	E	I	M	K	Q	T	W	Z	C	F	J	N
H	B	I	M	V	S	H	Q	T	P	W	Z	C	Y	E	F	J	N	L	R	U	X	A	D	G	K	O
I	C	J	N	W	T	I	R	U	Q	X	A	D	Z	F	G	K	O	M	S	V	Y	B	E	H	L	P
J	D	K	O	X	U	J	S	V	R	Y	B	E	A	G	H	L	P	N	T	W	Z	C	F	I	M	Q
K	E	L	P	Y	V	K	T	W	S	Z	C	F	B	H	I	M	Q	O	U	X	A	D	G	J	N	R
L	F	M	Q	Z	W	L	U	X	T	A	D	G	C	I	J	N	R	P	V	Y	B	E	H	K	O	S
M	G	N	R	A	X	M	V	Y	U	B	E	H	D	J	K	O	S	Q	W	Z	C	F	I	L	P	T
N	H	O	S	B	Y	N	W	Z	V	C	F	I	E	K	L	P	T	R	X	A	D	G	J	M	Q	U
O	I	P	T	C	Z	O	X	A	W	D	G	J	F	L	M	Q	U	S	Y	B	E	H	K	N	R	V
P	J	Q	U	D	A	P	Y	B	X	E	H	K	G	M	N	R	V	T	Z	C	F	I	L	O	S	W
Q	K	R	V	E	B	Q	Z	C	Y	F	I	L	H	N	O	S	W	U	A	D	G	J	M	P	T	X
R	L	S	W	F	C	R	A	D	Z	G	J	M	I	O	P	T	X	V	B	E	H	K	N	Q	U	Y
S	M	T	X	G	D	S	B	E	A	H	K	N	J	P	Q	U	Y	W	C	F	I	L	O	R	V	Z
T	N	U	Y	H	E	T	C	F	B	I	L	O	K	Q	R	V	Z	X	D	G	J	M	P	S	W	A
U	O	V	Z	I	F	U	D	G	C	J	M	P	L	R	S	W	A	Y	E	H	K	N	Q	T	X	B
V	P	W	A	J	G	V	E	H	D	K	N	Q	M	S	T	X	B	Z	F	I	L	O	R	U	Y	C
W	Q	X	B	K	H	W	F	I	E	L	O	R	N	T	U	Y	C	A	G	J	M	P	S	V	Z	D
X	R	Y	C	L	I	X	G	J	F	M	P	S	O	U	V	Z	D	B	H	K	N	Q	T	W	A	E
Y	S	Z	D	M	J	Y	H	K	G	N	Q	T	P	V	W	A	E	C	I	L	O	R	U	X	B	F
Z	T	A	E	N	K	Z	I	L	H	O	R	U	Q	W	X	B	F	D	J	M	P	S	V	Y	C	G

TABLE V

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/2}=\Theta_{p/1}$; $\Theta_{l/1}=\Theta_{o/2}$ ($\Theta_{l/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L
B	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P
C	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q
D	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J
E	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H
F	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B
G	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S
H	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T
I	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G
J	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U
K	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V
L	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W
M	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K
N	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O
O	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X
P	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y
Q	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z
R	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C
S	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E
T	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D
U	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M
V	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A
W	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N
X	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F
Y	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R
Z	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I

Table VI

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/2}=\Theta_{o/1}$; $\Theta_{1/1}=\Theta_{p/2}$ ($\Theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	T	P	G	J	U	L	I	M	F	C	Z	D	X	W	S	O	Q	K	H	E	B	Y	V	R	N
B	H	A	W	N	Q	B	S	P	T	M	J	G	K	E	D	Z	V	X	R	O	L	I	F	C	Y	U
C	L	E	A	R	U	F	W	T	X	Q	N	K	O	I	H	D	Z	B	V	S	P	M	J	G	C	Y
D	U	N	J	A	D	O	F	C	G	Z	W	T	X	R	Q	M	I	K	E	B	Y	V	S	P	L	H
E	R	K	G	X	A	L	C	Z	D	W	T	Q	U	O	N	J	F	H	B	Y	V	S	P	M	I	E
F	G	Z	V	M	P	A	R	O	S	L	I	F	J	D	C	Y	U	W	Q	N	K	H	E	B	X	T
G	P	I	E	V	Y	J	A	X	B	U	R	O	S	M	L	H	D	F	Z	W	T	Q	N	K	G	C
H	S	L	H	Y	B	M	D	A	E	X	U	R	V	P	O	K	G	I	C	Z	W	T	Q	N	J	F
I	O	H	D	U	X	I	Z	W	A	T	Q	N	R	L	K	G	C	E	Y	V	S	P	M	J	F	B
J	V	O	K	B	E	P	G	D	H	A	X	U	Y	S	R	N	J	L	F	C	Z	W	T	Q	M	I
K	Y	R	N	E	H	S	J	G	K	D	A	X	B	V	U	Q	M	O	I	F	C	Z	W	T	P	L
L	B	U	Q	H	K	V	M	J	N	G	D	A	E	Y	X	T	P	R	L	I	F	C	Z	W	S	O
M	X	Q	M	D	G	R	I	F	J	C	Z	W	A	U	T	P	L	N	H	E	B	Y	V	S	O	K
N	D	W	S	J	M	X	O	L	P	I	F	C	G	A	Z	V	R	T	N	K	H	E	B	Y	U	Q
O	E	X	T	K	N	Y	P	M	Q	J	G	D	H	B	A	W	S	U	O	L	I	F	C	Z	V	R
P	I	B	X	O	R	C	T	Q	U	N	K	H	L	F	E	A	W	Y	S	P	M	J	G	D	Z	V
Q	M	F	B	S	V	G	X	U	Y	R	O	L	P	J	I	E	A	C	W	T	Q	N	K	H	D	Z
R	K	D	Z	Q	T	E	V	S	W	P	M	J	N	H	G	C	Y	A	U	R	O	L	I	F	B	X
S	Q	J	F	W	Z	K	B	Y	C	V	S	P	T	N	M	I	E	G	A	X	U	R	O	L	H	D
T	T	M	I	Z	C	N	E	B	F	Y	V	S	W	Q	P	L	H	J	D	A	X	U	R	O	K	G
U	W	P	L	C	F	Q	H	E	I	B	Y	V	Z	T	S	O	K	M	G	D	A	X	U	R	N	J
V	Z	S	O	F	I	T	K	H	L	E	B	Y	C	W	V	R	N	P	J	G	D	A	X	U	Q	M
W	C	V	R	I	L	W	N	K	O	H	E	W	F	U	T	U	Q	S	M	J	G	D	V	X	T	P
X	F	Y	U	L	O	Z	Q	N	R	K	H	A	I	C	B	X	T	V	P	M	J	G	D	A	W	S
Y	J	C	Y	P	S	D	U	R	V	O	L	I	M	G	F	B	X	Z	T	Q	N	K	H	E	A	W
Z	N	G	C	T	W	H	Y	V	Z	S	P	M	Q	K	J	F	B	D	X	U	R	O	L	I	E	A

TABLE VII

Components:

(1)— A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2)— F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/2}=\Theta_{p/1}$; $\Theta_{1/2}=\Theta_{c/1}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
B	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
C	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
D	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
E	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
F	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
G	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
H	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
I	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
J	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
K	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
L	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
M	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
N	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
O	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
P	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Q	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
R	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
S	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
T	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
U	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
V	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
W	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
X	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
Y	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Z	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S

TABLE VIII

Componets:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/2}=\Theta_{c/1}$; $\Theta_{1/2}=\Theta_{p/1}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
D	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
E	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
F	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
G	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
H	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
K	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
L	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
M	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
N	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
O	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
P	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
Q	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
R	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
S	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
T	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
U	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
V	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
Y	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
Z	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G

TABLE IX [2]

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/1}=\Theta_{p/2}$; $\Theta_{1/1}=\Theta_{c/2}$ ($\Theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	V	F	R	T	S	X	I	E	Z	D	M	A	U	W	N	B	C	Y	G	H	J	K	L	O	P	Q
C	K	X	Y	H	G	O	Z	S	Q	T	U	V	J	L	W	F	R	P	I	E	D	M	A	N	B	C
D	M	O	P	E	I	N	Q	G	C	H	J	K	D	A	L	X	Y	B	Z	S	T	U	V	W	F	R
E	U	N	B	S	Z	W	C	I	R	E	D	M	T	V	A	O	P	F	Q	G	H	J	K	L	X	Y
F	J	W	F	G	Q	L	R	Z	Y	S	T	U	H	K	V	N	B	X	C	I	E	D	M	A	O	P
G	D	L	X	I	C	A	Y	Q	P	G	H	J	E	M	K	W	F	O	R	Z	S	T	U	V	N	B
H	T	A	O	Z	R	V	P	C	B	I	E	D	S	U	M	L	X	N	Y	Q	G	H	J	K	W	F
I	H	V	N	Q	Y	K	B	R	F	Z	S	T	G	J	U	A	O	W	P	C	I	E	D	M	L	X
J	E	K	W	C	P	M	F	Y	X	Q	G	H	I	D	J	V	N	L	B	R	Z	S	T	U	A	O
K	S	M	L	R	B	U	X	P	O	C	I	E	Z	T	D	K	W	A	F	Y	Q	G	H	J	V	N
L	G	U	A	Y	F	J	O	B	N	R	Z	S	Q	H	T	M	L	V	X	P	C	I	E	D	K	W
M	I	J	V	P	X	D	N	F	W	Y	Q	G	C	E	H	U	A	K	O	B	R	Z	S	T	M	L
N	Z	D	K	B	O	T	W	X	L	P	C	I	R	S	E	J	V	M	N	F	Y	Q	G	H	U	A
O	Q	T	M	F	N	H	L	O	A	B	R	Z	Y	G	S	D	K	U	W	X	P	C	I	E	J	V
P	C	H	U	X	W	E	A	N	V	F	Y	Q	P	I	G	T	M	J	L	O	B	R	Z	S	D	K
Q	R	E	J	O	L	S	V	W	K	X	P	C	B	Z	I	H	U	D	A	N	F	Y	Q	G	T	M
R	Y	S	D	N	A	G	K	L	M	O	B	R	F	Q	Z	E	J	T	V	W	X	P	C	I	H	U
S	P	G	T	W	V	I	M	A	U	N	F	Y	X	C	Q	S	D	H	K	L	O	B	R	Z	E	J
T	B	I	H	L	K	Z	U	V	J	W	X	P	O	R	C	G	T	E	M	A	N	F	Y	Q	S	D
U	F	Z	E	A	M	Q	J	K	D	L	O	B	N	Y	R	I	H	S	U	V	W	X	P	C	G	T
V	X	Q	S	V	U	C	D	M	T	A	N	F	W	P	Y	Z	E	G	J	K	L	O	B	R	I	H
W	O	C	G	K	J	R	T	U	H	V	W	X	L	B	P	Q	S	I	D	M	A	N	F	Y	Z	E
X	N	R	I	M	D	Y	H	J	E	K	L	O	A	F	B	C	G	Z	T	U	V	W	X	P	Q	S
Y	W	Y	Z	U	T	P	E	D	S	M	A	N	V	X	F	R	I	Q	H	J	K	L	O	B	C	G
Z	L	P	Q	J	H	B	S	T	G	U	V	W	K	O	X	Y	Z	C	E	D	M	A	N	F	R	I

[2] An interesting fact about this case is that if the plain component is made identical with the cipher component (both being the sequence FBPY ...), and if the enciphering equations are the same as for Table 1-B, then the resultant cipher square is identical with Table IX, except that the key letters at the left are in the order of the reversed mixed component, FXON .... In other words, the secondary cipher alphabets produced by the interaction of two identical mixed components are the same as those given by the interaction of a mixed component and the normal component.

TABLE X[3]

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/1}=\Theta_{c/2}$; $\Theta_{1/1}=\Theta_{p/2}$ ($\Theta_{1/1}$ is A).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	L	P	Q	J	H	B	S	T	G	U	V	W	K	O	X	Y	Z	C	E	D	M	A	N	F	R	I
C	W	Y	Z	U	T	P	E	D	S	M	A	N	V	X	F	R	I	Q	H	J	K	L	O	B	C	G
D	N	R	I	M	D	Y	H	J	E	K	L	O	A	F	B	C	G	Z	T	U	V	W	X	P	Q	S
E	O	C	G	K	J	R	T	U	H	V	W	X	L	B	P	Q	S	I	D	M	A	N	F	Y	Z	E
F	X	Q	S	V	U	C	D	M	T	A	N	F	W	P	Y	Z	E	G	J	K	L	O	B	R	I	H
G	F	Z	E	A	M	Q	J	K	D	L	O	B	N	Y	R	I	H	S	U	V	W	X	P	C	G	T
H	B	I	H	L	K	Z	U	V	J	W	X	P	O	R	C	G	T	E	M	A	N	F	Y	Q	S	D
I	P	G	T	W	V	I	M	A	U	N	F	Y	X	C	Q	S	D	H	K	L	O	B	R	Z	E	J
J	Y	S	D	N	A	G	K	L	M	O	B	R	F	Q	Z	E	J	T	V	W	X	P	C	I	H	U
K	R	E	J	O	L	S	V	W	K	X	P	C	B	Z	I	H	U	D	A	N	F	Y	Q	G	T	M
L	C	H	U	X	W	E	A	N	V	F	Y	Q	P	I	G	T	M	J	L	O	B	R	Z	S	D	K
M	Q	T	M	F	N	H	L	O	A	B	R	Z	Y	G	S	D	K	U	W	X	P	C	I	E	J	V
N	Z	D	K	B	O	T	W	X	L	P	C	I	R	S	E	J	V	M	N	F	Y	Q	G	H	U	A
O	I	J	V	P	X	D	N	F	W	Y	Q	G	C	E	H	U	A	K	O	B	R	Z	S	T	M	L
P	G	U	A	Y	F	J	O	B	N	R	Z	S	Q	H	T	M	L	V	X	P	C	I	E	D	K	W
Q	S	M	L	R	B	U	X	P	O	C	I	E	Z	T	D	K	W	A	F	Y	Q	G	H	J	V	N
R	E	K	W	C	P	M	F	Y	X	Q	G	H	I	D	J	V	N	L	B	R	Z	S	T	U	A	O
S	H	V	N	Q	Y	K	B	R	F	Z	S	T	G	J	U	A	O	W	P	C	I	E	D	M	L	X
T	T	A	O	Z	R	V	P	C	B	I	E	D	S	U	M	L	X	N	Y	Q	G	H	J	K	W	F
U	D	L	X	I	C	A	Y	Q	P	G	H	J	E	M	K	W	F	O	R	Z	S	T	U	V	N	B
V	J	W	F	G	Q	L	R	Z	Y	S	T	U	H	K	V	N	B	X	C	I	E	D	M	A	O	P
W	U	N	B	S	Z	W	C	I	R	E	D	M	T	V	A	O	P	F	Q	G	H	J	K	L	X	Y
X	M	O	P	E	I	N	Q	G	C	H	J	K	D	A	L	X	Y	B	Z	S	T	U	V	W	F	R
Y	K	X	Y	H	G	O	Z	S	Q	T	U	V	J	L	W	F	R	P	I	E	D	M	A	N	B	C
Z	V	F	R	T	S	X	I	E	Z	D	M	A	U	W	N	B	C	Y	G	H	J	K	L	O	P	Q

[3] Footnote 2 to Table IX, page 104, also applies to this table, except that the key letters at the left will follow the order of the direct mixed component.

TABLE XI

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/1}=\Theta_{p/2}$; $\Theta_{1/2}=\Theta_{c/1}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	G	Z	V	M	P	A	R	O	S	L	I	F	J	D	C	Y	U	W	Q	N	K	H	E	B	X	T
B	H	A	W	N	Q	B	S	P	T	M	J	G	K	E	D	Z	V	X	R	O	L	I	F	C	Y	U
C	I	B	X	O	R	C	T	Q	U	N	K	H	L	F	E	A	W	Y	S	P	M	J	G	D	Z	V
D	J	C	Y	P	S	D	U	R	V	O	L	I	M	G	F	B	X	Z	T	Q	N	K	H	E	A	W
E	K	D	Z	Q	T	E	V	S	W	P	M	J	N	H	G	C	Y	A	U	R	O	L	I	F	B	X
F	L	E	A	R	U	F	W	T	X	Q	N	K	O	I	H	D	Z	B	V	S	P	M	J	G	C	Y
G	M	F	B	S	V	G	X	U	Y	R	O	L	P	J	I	E	A	C	W	T	Q	N	K	H	D	Z
H	N	G	C	T	W	H	Y	V	Z	S	P	M	Q	K	J	F	B	D	X	U	R	O	L	I	E	A
I	O	H	D	U	X	I	Z	W	A	T	Q	N	R	L	K	G	C	E	Y	V	S	P	M	J	F	B
J	P	I	E	V	Y	J	A	X	B	U	R	O	S	M	L	H	D	F	Z	W	T	Q	N	K	G	C
K	Q	J	F	W	Z	K	B	Y	C	V	S	P	T	N	M	I	E	G	A	X	U	R	O	L	H	D
L	R	K	G	X	A	L	C	Z	D	W	T	Q	U	O	N	J	F	H	B	Y	V	S	P	M	I	E
M	S	L	H	Y	B	M	D	A	E	X	U	R	V	P	O	K	G	I	C	Z	W	T	Q	N	J	F
N	T	M	I	Z	C	N	E	B	F	Y	V	S	W	Q	P	L	H	J	D	A	X	U	R	O	K	G
O	U	N	J	A	D	O	F	C	G	Z	W	T	X	R	Q	M	I	K	E	B	Y	V	S	P	L	H
P	V	O	K	B	E	P	G	D	H	A	X	U	Y	S	R	N	J	L	F	C	Z	W	T	Q	M	I
Q	W	P	L	C	F	Q	H	E	I	B	Y	V	Z	T	S	O	K	M	G	D	A	X	U	R	N	J
R	X	Q	M	D	G	R	I	F	J	C	Z	W	A	U	T	P	L	N	H	E	B	Y	V	S	O	K
S	Y	R	N	E	H	S	J	G	K	D	A	X	B	V	U	Q	M	O	I	F	C	Z	W	T	P	L
T	Z	S	O	F	I	T	K	H	L	E	B	Y	C	W	V	R	N	P	J	G	D	A	X	U	Q	M
U	A	T	P	G	J	U	L	I	M	F	C	Z	D	X	W	S	O	Q	K	H	E	B	Y	V	R	N
V	B	U	Q	H	K	V	M	J	N	G	D	A	E	Y	X	T	P	R	L	I	F	C	Z	W	S	O
W	C	V	R	I	L	W	N	K	O	H	E	B	F	Z	Y	U	Q	S	M	J	G	D	A	X	T	P
X	D	W	S	J	M	X	O	L	P	I	F	C	G	A	Z	V	R	T	N	K	H	E	B	Y	U	Q
Y	E	X	T	K	N	Y	P	M	Q	J	G	D	H	B	A	W	S	U	O	L	I	F	C	Z	V	R
Z	F	Y	U	L	O	Z	Q	N	R	K	H	E	I	C	B	X	T	V	P	M	J	G	D	A	W	S

TABLE XII

Components:

(1) A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
(2) F B P Y R C Q Z I G S E H T D J U M K V A L W N O X

Enciphering equations: $\Theta_{k/1}=\Theta_{e/2}$; $\Theta_{1/2}=\Theta_{p/1}$ ($\Theta_{1/2}$ is F).

PLAIN TEXT

KEY	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B
B	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P
C	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y
D	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R
E	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C
F	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q
G	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z
H	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I
I	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G
J	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S
K	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E
L	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T	H
M	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D	T
N	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J	D
O	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U	J
P	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M	U
Q	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K	M
R	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V	K
S	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A	V
T	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L	A
U	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W	L
V	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N	W
W	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O	N
X	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X	O
Y	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F	X
Z	X	O	N	W	L	A	V	K	M	U	J	D	T	H	E	S	G	I	Z	Q	C	R	Y	P	B	F

APPENDIX 2[1]

Elementary Statistical Theory Applicable to the Phenomena of Repetition in Cryptanalysis

1. Introductory.—*a.* In Par. 9*c* it was stated that the phenomena of repetition in cryptanalytics may be removed from the realm of intuition and dealt with statistically. The discussion of the matter will here be confined to relatively simple phases of the theory of probability, a definition of which implies philosophical questions of no practical interest to the student of cryptanalysis. For his purposes, the following definition of *a priori* probability will be sufficient:

> The probability that an event will occur is the ratio of the number of "favorable cases" to the number of total possible cases, all cases being equally likely to occur. By a "favorable case" is meant one which will produce the event in question.

b. In what follows, reference will be made to *random assortments* of letters and especially to *random text*. By the latter will be meant merely that the text under consideration has been assumed to have been enciphered by some more or less complex cryptographic system so that for all practical purposes the sequence of letters constituting this text is a random assortment; that is, the sequence is just about what would have been obtained if the letters had been drawn at random out of a box containing a large number of the 26 letters of the alphabet, all in equal proportions, so that there are exactly the same numbers of A's, B's, C's, . . . Z's. It is assumed that each time in making a drawing from such a box, the latter is thoroughly shaken so that the letters are thoroughly mixed and then a single letter is selected at random, recorded, and replaced in the same box. In what follows, the word "box" will refer to the box as described.

c. A uniliteral frequency distribution of a large volume of random text will be "flat," i. e., lacking crests and troughs.

d. For purposes of statistical analysis, the text of a monoalphabetic substitution cipher is equivalent to plain text. As a corollary, when a polyalphabetic substitution cipher has been reduced to the simple terms of a set of monoalphabets, i. e., when the letters constituting the cipher text have been allocated into their proper uniliteral distributions, the letters falling into the respective distributions are statistically equivalent to plain text.

2. Data pertaining to single letters.—*a.* (1) A single letter will be drawn at random from the box. What is the probability that it will be an A? According to the foregoing definition of probability, since the total number of possible cases is 26 and the number of favorable cases is here only 1, the probability is $1:26=\frac{1}{26}=.0385$. This is the probability of drawing an A from the box. The probability that the letter drawn will be a B, a C, a D, . . ., a Z is the same as for A. In other words, the probability of drawing *any specified single letter* is $p=.0385$.

(2) The value $p=.0385$, as found above, may also be termed the probability constant for single letters in random text of a 26-letter alphabet. For any language this constant is merely the reciprocal of the total number of different characters which may be employed in writing the text in question.

[1] In the preparation of this appendix, the author has had the benefit of the very helpful suggestions of Capt. H. G. Miller, Signal Corps, Mr. F. B. Rowlett, Dr. S. Kullback, and Dr. A. Sinkov, Assistant Cryptanalysts, O. C. Sig. O. Certain parts of Dr. Kullback's important paper "Statistical Methods in Cryptanalysis" form the basis of the discussion.

(3) Another way of interpreting the notation $p=.0385$ is to say that in a large volume of random text, for example in 100,000 letters, any letter that one may choose to specify may be expected to occur about 3,850 times; in 10,000 letters it may be expected to occur about 385 times; in 1,000 letters, about 38.5 times, and so on. In every-day language it would be said that "in the long run" or "on the average" in 1,000 letters of random text there will be about 38.5 occurrences of each of the 26 letters of the alphabet.

(4) But unfortunately, in cryptanalysis it is not often the case that one has such a large number of letters available for study in any single cipher alphabet. More often the cryptanalyst has a relatively small number of letters and these must be distributed over several cipher alphabets. Hence it is necessary to be able to deal with smaller numbers of letters. Consider a specific piece of random text of only 100 letters. It has been seen that "in the long run" each letter may be expected to occur about 3.85 times in this amount of random text; that is, the 26 letters will have an *average* frequency of 3.85. But in reaching this average of 3.85 occurrences in 100 letters, it is obvious that some letter or letters may not appear at all, some may appear once, some twice, and so on. How many will not appear at all; how many will appear 1, 2, 3, . . . times? In other words, how will the different categories of letters (different in respect to frequency of occurrence) be distributed, or what will the *distribution* be like? Will it follow any kind of law or pattern? The cryptanalyst also wants to know the answer to questions such as these: What is the probability that a specified letter will not appear at all in a given piece of text? That it will appear *exactly* 1, 2, 3, . . . times? That it will appear *at least* 1, 2, 3, . . . times? The same sort of questions may be asked with respect to digraphs, trigraphs, and so on.

b. (1) It may be stated at once that questions of this nature are not easily answered, and a complete discussion falls quite outside the scope of this text. However, it will be sufficient for the present purposes if the student is provided with a more or less simple and practical means of finding the answers. With this in view certain curves have been prepared from data based upon Poisson's exponential expansion, or the "law of small probabilities" and their use will now be explained. Students without a knowledge of the mathematical theory of probability and statistics will have to take the curves "on faith" Those interested in their derivation are referred to the following texts:

Fisher, R. A., *Statistical Methods for Research Workers*, London, 1937.
Fry, T. C., *Probability and Its Engineering Uses*, New York, 1928.

(2) By means of these *probability curves*, it is possible to find, in a relatively easy manner, the probability for 0, 1, 2, . . . 11 occurrences of an event in n cases, if the *mean* (expected, average, probable) number of occurrences in these n cases is known. For example, given a cryptogram equivalent to 100 letters of random text, what is the probability that any specified single letter, whatever will not appear at all in the cryptogram? Since the probability of the occurrence of a specified single letter is $\frac{1}{26}=.0385$, and there are 100 letters in the cryptogram, the average or expected or mean number of occurrences of an A, a B, a C, . . ., is $.0385\times100=3.85$. Refer now to that probability curve which is marked "f_0", meaning "frequency zero", or "zero occurrences." On the horizontal or x axis of that curve find the point corresponding to the value 3.85 and follow the vertical coordinate determined by this value up to the point of intersection with the curve itself; then follow the horizontal coordinate determined by this intersection point over to the left and read the value on the vertical axis of the curve. It is approximately .021. This means that the probability that a specified single letter (an A, a B, a C, . . .) will not appear at all in the cryptogram, if it really were a perfectly random assortment of 100 letters, is .021.

152018—38——8

That is, according to the theory of probability, in 1,000 cases of random-text messages of 100 letters each, one may expect to find about 21 messages in which a specified single letter will not appear at all. Another way of saying the same thing is: If 1,000 sets of 100 letters of random text are examined, in about 21 out of the 1,000 such sets any letter that one may choose to name will be absent. This, of course, is merely a theoretical expectancy; it indicates only what probably will happen in the long run.

(3) What is the probability that a specified single letter will appear *exactly* once in 100 letters of random text? To answer this question, find on the curve marked f_1, the point of intersection of the vertical coordinate corresponding to the mean or average value 3.85 with the curve; follow the horizontal coordinate thus determined over to the vertical scale at the left; read the value on this scale. It is .082, which means that in 1,000 cases of random-text messages of 100 letters each, one may expect to find about 82 messages in which any letter one chooses to specify will occur exactly once, no more and no less.

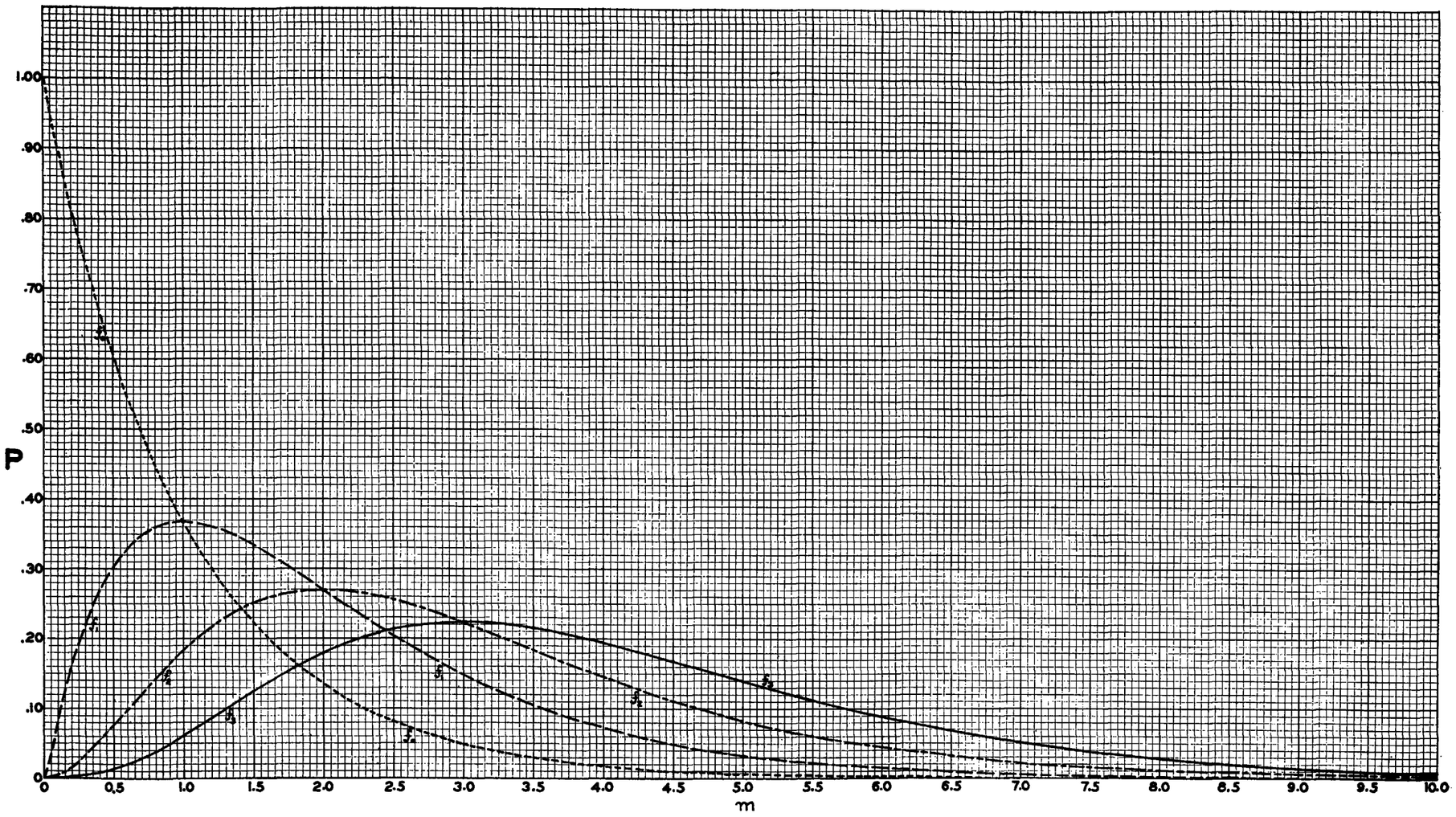
(4) In the same way, the probability that a specified single letter will appear *exactly* twice is found to be .158; exactly 3 times, .202; and so on, as shown in the table below:

100 letters of random text

Frequency (x)	Probability that a specified single letter will occur exactly x times
0	0. 021
1	. 082
2	. 158
3	. 202
4	. 195
5	. 150
6	. 096
7	. 053
8	. 026
9	. 011
10	. 004
11	. 001

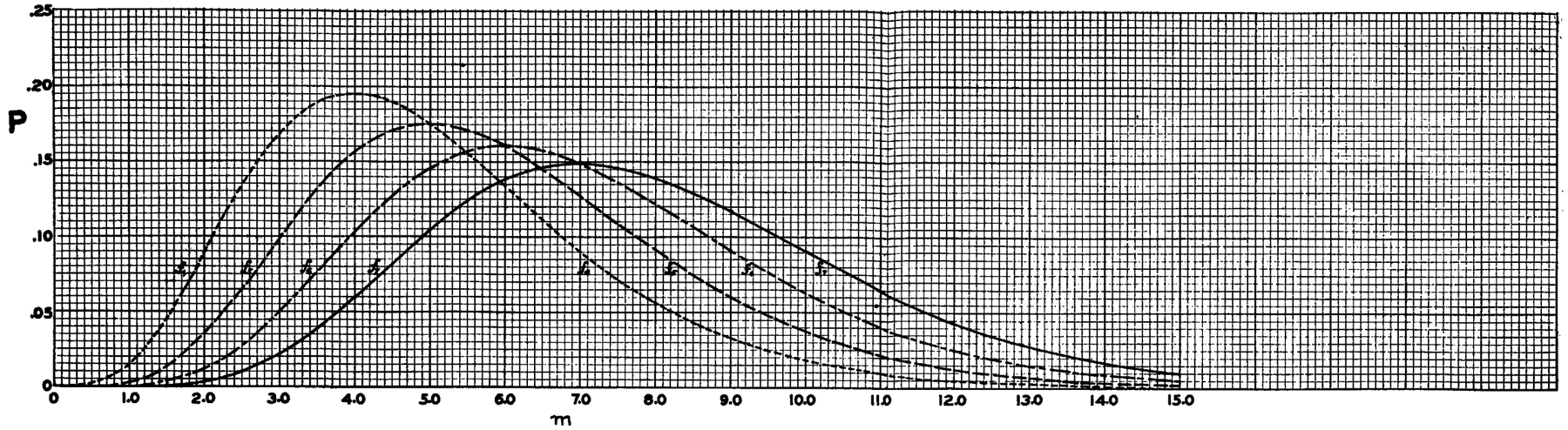
(5) To find the probability that a specified single letter will occur *at least* 1, 2, 3, . . . times in a series of letters constituting random text, one reasons as follows: Since the concept "at least 1" implies that the number specified is to be considered only as the minimum, with no limit indicated as to maximum, occurrences of 2, 3, 4, . . . are also "favorable" cases; the probabilities for *exactly* 1, 2, 3, 4, . . . occurrences should therefore be added and this will give the probability for "at least 1." Thus, in the case of 100 letters, the sum of the probabilities for exactly 1 to 11 occurrences, as set forth in the table directly above, is .978, and the latter value approximates the probability for at least 1 occurrence.

(6) A more accurate result will be obtained by the following reasoning. The probability for zero occurrences is .021. Since it is certain that a specified letter will occur either zero times or 1, 2, 3, . . . times, to find the probability for *at least* one time it is merely necessary to subtract the probability for zero occurrences from unity. That is, $1-.021=.979$, which is .001 greater than the result obtained by the other method. The reason it is greater is that the value .979 includes occurrences beyond 11, which were excluded from the previous calculation. Of course, the probabilities for these occurrences beyond 11 are very small, but taken all together they

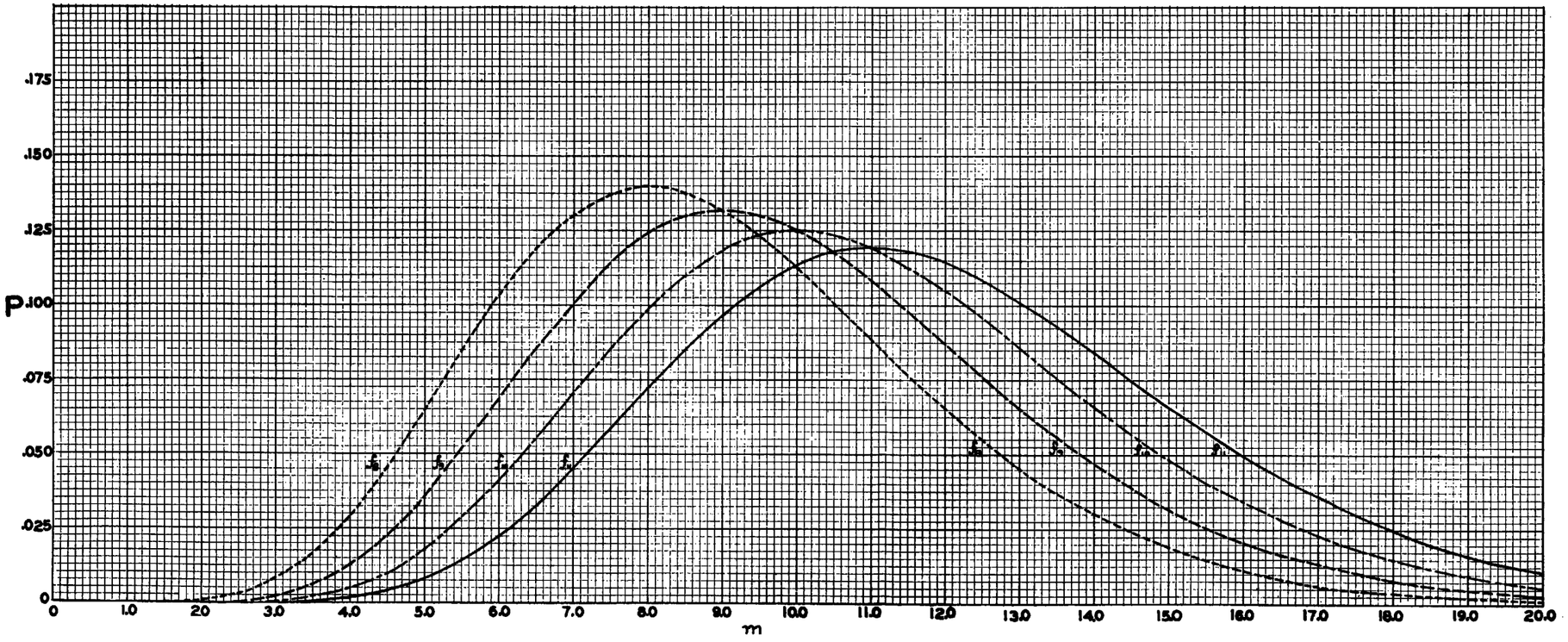


Curves showing probability for 0, 1, 2, and 3 occurrences of an event in n cases, given the mean number of occurrences.

152018—38 (Face p. 110) No. 1



Curves showing probability for 4, 5, 6, and 7 occurrences of an event in n cases, given the mean number of occurrences.



Curves showing probability for 8, 9, 10, and 11 occurrences of an event in n cases, given the mean number of occurrences.

152018—38 (Face p. 110) No. 2

add up to .001, the difference between the results obtained by the two methods. The probability for at least 2 occurrences is the difference between unity and the sum of the probability for zero and exactly 1 occurrences; that is, $1-(P_0+P_1)=1-(.021+.082)=1-.103=.897$. The respective probabilities for various numbers of occurrences of a specified single letter (from 0 to 11) are given in the following table:

100 letters of random text

Frequency (x)	Probability that a specified single letter will occur exactly x times	Probability that a specified single letter will occur at least x times
0	0.021	1.000
1	.082	.979
2	.158	.897
3	.202	.739
4	.195	.537
5	.150	.342
6	.096	.192
7	.053	.096
8	.026	.043
9	.011	.017
10	.004	.006
11	.001	.002

(7) The foregoing calculations refer to random text composed of 100 letters. For other numbers of letters, it is merely necessary to find the mean (multiply the probability for drawing a specified single letter out of the box, which is $\frac{1}{26}$ or .0385, by the number of letters in the assortment) and refer to the various curves, as before. For example, for a random assortment of 200 letters, the mean is $200 \times .0385$, or 7.7, and this is the value of the point to be sought along the horizontal or x axes of the curves; the intersections of the respective vertical lines corresponding to this mean with the various curves for 0, 1, 2, 3, . . . occurrences give the probabilities for these occurrences, the reading being taken on the vertical or y axes of the curves.

(8) The discussion thus far has dealt with the probabilities for 0, 1, 2, 3, . . . occurrences of specified single letters. It may be of more practical advantage to the student if he could be shown how to find the answer to these questions: Given a random assortment of 100 letters *how many* letters may be expected to occur *exactly* 0, 1, 2, 3, . . . times? How many may be expected to occur *at least* 1, 2, 3, . . . times? The curves may here again be used to answer these questions, by a very simple calculation: multiply the probability value as obtained above for a specified single letter by the number of different elements being considered. For example, the probability that a specified single letter will occur exactly twice in a perfectly random assortment of 100 letters is .158; since the number of different letters is 26, the absolute number of single letters that may be expected to occur exactly 2 times in this assortment is $.158 \times 26 = 4.108$. That is, in 100 letters of randm text there should be about four letters which occur exactly 2 times. The following table gives the data for various numbers of occurrences.

100 letters of random text

Frequency (x)	Probability that a specified single letter will occur exactly x times	Probability that a specified single letter will occur at least x times	Probable number of letters appearing exactly x times	Probable number of letters appearing at least x times
0	0.021	1.000	0.546	26.000
1	.082	.979	2.132	25.454
2	.158	.897	4.108	23.322
3	.202	.739	5.252	19.214
4	.195	.537	5.070	13.962
5	.150	.342	3.900	8.892
6	.096	.192	2.496	4.992
7	.053	.096	1.378	2.496
8	.026	.043	.676	1.118
9	.011	.017	.286	.442
10	.004	.006	.104	.156
11	.001	.002	.026	.052

(9) Referring again to the curves, and specifically to the tabulated results set forth directly above, it will be seen that the probability that there will be exactly two occurrences of a specified single letter in 100 letters of random text (.158), is less than the probability that there will be exactly three occurrences (.202); in other words, the chances that a specified single letter will occur exactly three times are better, by about 25 percent, than that it will occur only two times. Furthermore, there will be about five letters which will occur exactly 3 times, and about five which will occur exactly 4 times, whereas there will be only about two letters which will occur exactly 1 time. Other facts of a similar import may be deduced from the foregoing table.

c. The discussion thus far has dealt with random assortments of letters. What about other types of texts, for example, normal plain text? What is the probability that E will occur 0, 1, 2, 3, . . . times in 50 letters of normal English? The relative frequency value or probability that a letter selected at random from a large volume of normal English text will be E is .12604. (In 100,000 letters E occurred 12,604 times.) For 50 letters this value must be multiplied by 50, giving 6.3 as the mean or point to be found along the x axes of the curves. The probabilities for 0, 1, 2, 3, . . . occurrences are tabulated below:

50 letters of normal English plain text

Frequency (x)	Probability that an E will be drawn exactly x times	Probability that an E will be drawn at least x times
0	0.002	1.000
1	.011	.998
2	.036	.987
3	.076	.951
4	.120	.875
5	.151	.755
6	.159	.604
7	.143	.445
8	.113	.302
9	.079	.223
10	.050	.173
11	.029	.123

d. (1) It has been seen that the probability of occurrence of a specified single letter in random text employing a 26-letter alphabet is $p=\frac{1}{26}=.0385$. If a considerable volume of such text is written on a large sheet of paper and a pencil is directed at random toward this text, the probability that the pencil point will hit the letter A, *or any other letter which may be specified in advance,* is .0385. Now suppose two pencils are directed simultaneously toward the sheet of paper. The probability that both pencil points will hit two A's is $\frac{1}{26}\times\frac{1}{26}=\frac{1}{26^2}=.00148$, since in this case one is dealing with the probability of the simultaneous occurrence of two events which are independent. The probability of hitting two B's, two C's, . . ., two Z's is likewise $\frac{1}{26^2}$. Hence, if no particular letter is specified, and merely this question is asked: "What is the probability that both pencil points will hit the same letter?" the answer must be the sum of the separate probabilities for simultaneously hitting two A's, two B's, and so on, for the whole alphabet, which is $26\times\frac{1}{26^2}=\frac{1}{26}=.0385$. This, then, is the probability that *any* two letters selected *at random* in random text of a 26-letter alphabet will be identical or will *coincide.* Since this value remains the same so long as the number of alphabetic elements remains fixed, it may be said that *the probability of monographic coincidence in random text of a 26-element alphabet is* .0385. The foregoing italicized expression [2] is important enough to warrant assigning a special symbol to it, *viz,* κ_r (read "kappa sub-*r*"). For a 26-element alphabet, then, $\kappa_r=.0385$.

(2) Now if one asks: "Given a random assortment of 10 letters, what are the respective probabilities of occurrence of 0, 1, 2, . . . single-letter coincidences?" one proceeds as follows. As before, it is first necessary to find the mean or expected number of coincidences and then refer to the various probability curves. To find the mean, one reasons as follows. Given a sequence of 10 letters, one may begin with the 1st letter and compare it with the 2d, 3d, . . . 10th letter to see if any two letters coincide; 9 such comparisons may be made, or in other words there are, beginning with the 1st letter, 9 opportunities for the occurrence of a coincidence. But one may also start with the 2nd letter and compare it with the 3d, 4th . . . 10th letter, thus yielding 8 more opportunities for the occurrence of a coincidence, and so on. This process may continue until one reaches the 9th letter and compares it with the 10th, yielding but one opportunity for the occurrence in question. The total number of comparisons that can be made is therefore the sum of the series of numbers 9, 8, 7, . . . 1, which is 45 comparisons.[3] Since in the 10 letters there are 45 opportunities for coincidence of single letters, and since the probability

---

[2] The expression itself may be termed a *parameter,* which in mathematics is often used to designate a constant that characterizes by each of its particular values some particular member of a system of values, functions, etc. The word is applicable in the case under discussion because the value obtained for κ_r is .0385; for a 25-element alphabet, $\kappa_r=.0400$; for a 27-element alphabet, $\kappa_r=.0370$, etc.

[3] The number of comparisons may readily be found by the formula $\frac{n(n-1)}{2}$, where n is the total number of letters involved. This formula is merely a special case under the general formula for ascertaining the number of combinations that may be made of n different things taken r at a time, which is ${}_nC_r=\frac{n!}{r!\,(n-r)!}$. In the present case, since only two letters are compared at a time, r is always 2, and hence the expression $\frac{n!}{r!\,(n-r)!}$, which is the same as $\frac{n(n-1)\,(n-2)!}{2(n-2)!}$, becomes by cancellation of the term $(n-2)!$ reduced to $\frac{n(n-1)}{2}$.

for monographic coincidence in random text is .0385 the expected number of coincidences is .0385×45=1.7325. With m=1.7 one consults the various probability curves and an approximate distribution for exactly and for at least 0, 1, 2, . . . coincidences may readily be ascertained.[4]

e. (1) Now consider the matter of monographic coincidence in English plain text.[5] Following the same reasoning outlined in subpar. *d* (1), the probability of coincidence of two A's in plain text is the square of the probability of occurrence of the single letter A in such text. The probability of coincidence of two B's is the square of the probability of occurrence of the single letter B, and so on. The sum of these squares for all the letters of the alphabet, as shown in the following table, is found to be .0667.

Letter	Frequency [1] in 1,000 letters	Probability of separate occurrence of the letter	Square of probability of separate occurrence
A	73. 66	0. 0737	0. 0054
B	9. 74	. 0097	. 0001
C	30. 68	. 0307	. 0009
D	42. 44	. 0424	. 0018
E	129. 96	. 1300	. 0169
F	28. 32	. 0283	. 0008
G	16. 38	. 0164	. 0003
H	33. 88	. 0339	. 0012
I	73. 52	. 0735	. 0054
J	1. 64	. 0016	. 0000
K	2. 96	. 0030	. 0000
L	36. 42	. 0364	. 0013
M	24. 74	. 0247	. 0006
N	79. 50	. 0795	. 0063
O	75. 28	. 0753	. 0057
P	26. 70	. 0267	. 0007
Q	3. 50	. 0035	. 0000
R	75. 76	. 0758	. 0057
S	61. 16	. 0612	. 0037
T	91. 90	. 0919	. 0084
U	26. 00	. 0260	. 0007
V	15. 32	. 0153	. 0002
W	15. 60	. 0156	. 0002
X	4. 62	. 0046	. 0000
Y	19. 34	. 0193	. 0004
Z	. 98	. 0010	. 0000
Total	1,000. 00	1. 0000	. 0667

[1] The data given are taken from Table 3, Appendix 1, Military Cryptanalysis, Part I.

This then is the probability that any two letters selected at random in a large volume of normal English telegraphic plain text will coincide. Since this value remains the same so long as the character of the language does not change radically, it may be said that *the probability of monographic coincidence in English telegraphic plain text is* .0667, or κ_p=.0667.

[4] The approximation given by the Poisson distribution in the case of single letters is not as good as that in the case of digraphs, trigraphs, etc., discussed in paragraphs 3, 4, below.

[5] The theory of monographic coincidence in plain text was originally developed and applied by the author in a technical paper written in 1925 dealing with his solution of messages enciphered by a cryptograph known as the "Hebern Electric Super-Code." The paper was printed in 1934.

(2) Given 10 letters of English plain text, what is the probability that there will be 0, 1, 2, . . . single-letter coincidences? Following the line of reasoning in subparagraph *d* (2), the expected number of coincidences is $.0667 \times 45 = 3.00$, or $m = 3$. The distribution for exactly and for at least 0, 1, 2, . . . coincidences may readily be found by reference to the various probability curves. (See *footnote* 4.)

f. The fact that κ_p (for English) is almost twice as great as κ_r is of considerable importance in cryptanalysis. It will be dealt with in detail in a subsequent text. At this point it will merely be said that κ_p and κ_r for other languages and alphabets have been calculated and show considerable variation, as will be noted in the table shown in paragraph 3*d*.

3. Data pertaining to digraphs.—*a*. (1) The foregoing discussion has been restricted to questions concerning single letters, but by slight modification it can be applied to questions concerning digraphs, trigraphs, and longer polygraphs.

(2) In the preceding cases it was necessary, before referring to the various probability curves, to find the mean or expected number of occurrences of the event in question in the total number of cases or trials being considered. Given a piece of random text totalling 100 letters, for example, what is the mean (average, probable, expected) number of occurrences of digraphs in this text? Since there are 676 different digraphs, the probability of occurrence of any specified digraph is $\frac{1}{676} = .00148$; since in 100 letters there are 99 digraphs (if the letters are taken consecutively in pairs) the mean or average number of occurrences in this case is $.00148 \times 99 = .147$. Having the mean number of occurrences of the event under consideration, one may now find the answers to these questions: What is the probability that any specified digraph, say XY, will not occur? What is the probability that it will occur *exactly* 1, 2, 3, . . . times? *At least* 1, 2, 3, . . . times?

(3) Again the probability curves may be used as before, for the type of distribution is the same. The following values are obtainable by reference to the various curves, using the mean value $.00148 \times 99 = .147$.

100 letters of random text

Frequency (x)	Probability that a specified digraph will occur exactly x times	Probability that a specified digraph will occur at least x times	Probable number of digraphs appearing exactly x times	Probable number of digraphs appearing at least x times
0	0.86	1.00	581.36	676.00
1	.13	.14	87.88	94.64
2	.01	.01	6.76	6.76
3	.00	.00	0.00	0.00

(4) Thus it is seen that in 100 letters of random text the probability that a specified digraph will occur exactly once, for example, is .13; at least once, .14; at least twice, .01. The probability that a specified digraph will occur at least 3 times is negligible. (By calculation, it is found to to be .0005.)

b. (1) The probability of digraphic coincidence in random text based upon a 26-element alphabet is of course quite simply obtained: since there are 26^2 different digraphs, the probability of selecting any specified digraph in random text is $\frac{1}{26^2}$. The probability of selecting two identical digraphs in such text, *when the digraphs are specified*, is $\frac{1}{26^2} \times \frac{1}{26^2} = \frac{1}{26^4}$. Since there are 26^2 different digraphs, the probability of digraphic coincidence in random text, κ_r^2, is $26^2 \times \frac{1}{26^4} = \frac{1}{26^2} = .00148$.

(2) Given a random assortment of 100 letters, what is the probability of occurrence of 0, 1, 2, . . . digraphic coincidences? Following the line of reasoning in paragraph 2*d* (2), in 100 letters the total number of comparisons that may be made to see if two digraphs coincide is 4,851. This number is obtained as follows: Consider the 1st and 2d letters in the series of 100 letters; they may be combined to from a digraph to be compared with the digraphs formed by combining the 2d and 3d, the 3d and 4th, the 4th and 5th letters, and so on, giving a total of 98 comparisons. Consider the digraph formed by combining the 2d and 3d letters; it may be compared with the digraphs formed by combining the 3d and 4th, 4th and 5th letters, and so on, giving a total of 97 comparisons. This process may be continued down to the digraph formed by combining the 98th and 99th letters, which yields only one comparison, since it may be compared only with the digraph resulting from combining the 99th and 100th letters. The total number of comparisons is the sum of the sequence of numbers 98, 97, 96, 95, . . . 1, which is 4,851.[6]

(3) Since in the 100 letters there are 4,851 opportunities for the occurrence of a digraphic coincidence, and since $\kappa_r^2 = .00148$, the expected number of coincidences is $.00148 \times 4851 = 7.17948 = 7.2$. The various probability curves may now be referred to and the following results are obtained:

Distribution for 100 letters of random text

Frequency (x)	Probability for exactly x digraphic coincidences	Probability for at least x digraphic coincidences
0	0. 001	1. 000
1	. 005	. 999
2	. 019	. 994
3	. 046	. 975
4	. 083	. 929
5	. 120	. 846
6	. 144	. 726
7	. 148	. 582
8	. 134	. 434
9	. 107	. 300
10	. 077	. 193
11	. 050	. 116

c. In this table it will be noted that it is almost certain that in 100 letters of random text there will be at least one digraphic coincidence, despite the fact that there are 676 possible digraphs and only 99 of them have appeared in 100 letters. When one thinks of a total of 676 different digraphs from which the 99 digraphs may be selected it may appear rather incredible that the chances are better than even (.582) that one will find at least 7 digraphic coincidences in 100 letters of random text, yet that is what the statistical analysis of the problem shows to be the case. *These are, of course, purely accidental repetitions.* It is important that the student should fully realize that more coincidences or accidental repetitions than he feels intuitively should occur in random text will actually occur in the cryptograms he will study. He must therefore be on guard against putting too much reliance upon the surface appearances of the phenomena of repetition; he must calculate what may be expected from pure chance, to make sure that the number and length of the repetitions he does see in a cryptogram are really better than what may be expected in random text. In studying cryptograms composed of figures this

[6] The formula for finding the number of comparisons that can be made is as follows, where n=the total number of letters in the sequence and t is the length of the polygraph; No. of comparisons $= \frac{(n-t)(n-t+1)}{2}$.

is very important, for as the number of different symbols decreases the probability for purely chance coincidences increases.

d. (1) For convenience the following values of the reciprocals of various numbers from 20 to 36, and of the reciprocals of the squares, cubes, and 4th powers of these numbers are listed:

x	$1/x$	$1/x^2$	$1/x^3$	$1/x^4$
20	0. 0500	0. 002500	0. 000125	0. 00000625
21	. 0476	. 002266	. 000108	. 00000514
22	. 0455	. 002070	. 000094	. 00000429
23	. 0435	. 001892	. 000082	. 00000358
24	. 0417	. 001739	. 000073	. 00000302
25	. 0400	. 001600	. 000064	. 00000256
26	. 0385	. 001482	. 000057	. 00000220
27	. 0370	. 001369	. 000051	. 00000187
28	. 0357	. 001274	. 000046	. 00000162
29	. 0345	. 001190	. 000041	. 00000142
30	. 0333	. 001109	. 000037	. 00000123
31	. 0323	. 001043	. 000034	. 00000109
32	. 0313	. 000980	. 000031	. 00000096
33	. 0303	. 000918	. 000028	. 00000084
34	. 0294	. 000864	. 000025	. 00000075
35	. 0286	. 000818	. 000023	. 00000067
36	. 0278	. 000773	. 000021	. 00000060

(2) The following table gives the probabilities for monographic and digraphic coincidence for plain-text in several languages.

Language	κ_p	κ_{p^2}
English	0. 0667	0. 0069
French	. 0778	. 0093
German	. 0762	. 0112
Italian	. 0738	. 0081
Spanish	. 0775	. 0093

4. **Data pertaining to trigraphs, etc.**—*a.* Enough has been shown to make clear to the student how to calculate probability data concerning trigraphs, tetragraphs, and longer polygraphs.

b. (1) For example, in 100 letters of random text the value of m (the mean) for trigraphs is $.00005689 \times 100 = .005689$. With so small a value, the probability curves are hardly usable, but at any rate they show that the probability of occurrence of a specified trigraph in so small a volume of text is so small as to be practically negligible. The probability of a specified trigraph occurring twice in that text is an even smaller quantity.

(2) The calculation for finding the probability of at least one trigraphic coincidence in 100 letters of random text is as follows:

$$m = \left(\frac{97 \times 98}{2}\right)\left(\frac{1}{26^3}\right) = 4{,}753 \times .0000568912 = .2704 = .27$$

Referring to curve f_0, with $m = .27$ the probability of finding no trigraphic coincidence is .76. The probability of finding at least one trigraphic coincidence is therefore $1 - .76 = .24$.

c. The calculation for a tetragraphic coincidence is as follows:

$$m = \left(\frac{96 \times 97}{2}\right)\left(\frac{1}{26^4}\right) = 4{,}656 \times .0000021883 = .0101 = .01$$

Referring to curve f_0, with $m = .01$ the probability of finding no tetragraphic coincidence is so high as to amount almost to certainty. Consequently, the probability of finding at least

one tetragraphic coincidence is practically nil. (It is calculated to be .0094=approximately .01. This means that in a hundred cases of 100-letter random-text cryptograms, one might expect to find but one cryptogram in which a 4-letter repetition is brought about purely by chance; it is, in common parlance, a "hundred to one shot.") Consequently, if a tetragraphic repetition is found in a cryptogram of 100 letters, the probability that it is an accidental repetition is extremely small. If not accidental, then it must be causal, and the cause should be ascertained.

5. An example.—*a.* The message of Par. 9*a* of the text proper will be employed. First, let the repetitions be sought and underlined; then the repetitions are listed for convenience.

A. U S Y E S E C P M P L C C L N X B W C S O X U V D
B. S C R H T H X I P L I B C I J U S Y E E G U R D P
C. A Y B C X O F P J W J E M G P X V E U E L E J Y Q
D. M U S C X J Y M S G L L E T A L E D E C G B M F I

Group	Number of occurrences
BC	2
CX	2
EC	2
LE	3
JY	2
PL	2
SC	2
SY	2
US	3
YE	2
SYE	2
USY	2
USYE	2

b. Referring to the table in Par. 3*a* (3) above, it will be seen that in 100 letters of random text one might expect to find about 7 digraphs appearing at least twice and no digraph appearing 3 times. The list of repetitions shows 8 digraphs occurring twice and 2 occurring 3 times.

c. Again, the list of repetitions shows 10 digraphs each repeated at least twice; the table in Par. 3*b* (3) above shows that in 100 letters of random text the probability of finding at least that many digraphic coincidences is only .193. That is, the chances of this being an accident are but 176 in a thousand; or another way of expressing the same thing is to say that the odds against this phenomenon being an accident are as 807 is to 193 or roughly 4 to 1.

d. The probability of finding at least one trigraphic coincidence in 100 letters of random text is very small, as noted in Par. 4*b;* the probability of finding at least one tetragraphic coincidence is still smaller (Par. 4*c*). Yet this cipher message of but 100 letters contains a repetition of this length.

e. A consideration of the foregoing leads to the conclusion that the number and length of the repetitions manifested by the cryptogram are not accidental, such as might be expected to occur in random text of the same length; hence they must be causal in their origin. The cause in this case is not difficult to find: repeated isolated letters and repeated sequences of letters (digraphs, trigraphs) in the plain text were actually enciphered by identical alphabets, resulting in producing repeated letters and sequences in the cipher text.

INDEX

	Page
Accidental repetitions	12
Alphabets:	
Classification of	4
Derived	4
Interrelated	24
Mixed	24
Secondary	4
Analytical key	95
Aperiodic systems	2
Assumptions for values, check	76
Average	109
Bazeries	23
Beaufort	9, 19
Causal repetitions	12
Cipher disks	5
Classification of alphabets	4
Coincidence:	
Digraphic	115–116
Monographic	113, 114
Tetragraphic	117
Trigraphic	117–118
Comparisons, number of	113, 116
Completing the plain component sequence	19, 79, 82, 88
Component monoalphabets	15
Constant intervals	85
Conversion:	
Into monoalphabetic terms	92, 94
Into plain-component equivalents	81, 83
Cryptograms:	
In different keys, containing identical plain text	85
With plain text	84
Cyclic phenomena	2
Data pertaining to:	
Digraphs	115–116
Trigraphs	117–118
Decimation	Footnote 1 53
Delastelle	9
Derived alphabets	4
Digraphic coincidence	115–116
Probability for	115
Digraphs, data pertaining to	115
Direct symmetry	9, 26
Application of principles	32
Of position	84
Distribution of different categories of letters in respect to frequency of occurrence	109
Double-key system	Footnote 2 3
Equations, enciphering	5, 6
Equivalent primary components	53
Expected number of occurrences	109
Factoring	15
Intervals	86
Givierge	Footnote 1 23
Gronsfeld	21
Identical messages enciphered by keywords of different lengths	89*ff*
Identical superimpositions	86
Index letter	6
Indirect symmetry	9, 52
Of position	60, 68, 69, 84, 85
Interrelated alphabets	24
Latent symmetry	52
Law of small probabilities	109
Matching	47
Distributions	94
Mean number	109
Coincidences, of	114
Digraphs, of	115
Mixed alphabets	24
Monoalphabetic terms, conversion into	46
Monographic coincidence	113*ff*
Multiple alphabet system	3
Number of comparisons	113–116
Parameter	Footnote 2 113
Partial chains of equivalents	85
Period, determination of	10, 15
Periodic systems	2
Primary classification	3
Phenomena of repetition	108
Poe, Edgar Allan	Footnote 1 2
Poisson's exponential expansion	109
Polyalphabetic substitution:	
Distinguished from monoalphabetic	1
Primary classification	2
Sequence of study	3
Primary components	4, 5
Equivalent	53
Reconstruction of	27, 52
Principles of indirect symmetry of position:	
Application of principles	69*ff*
Application to specific example	60*ff*
Fundamental theory	68, 69
Probability:	
Definition of *apriori*	108
Of digraphic coincidence	115–116

	Page
Probability—Continued.	
Of monographic coincidence	113ff
Of tetragraphic coincidence	117
Of trigraphic coincidence	117
Probable-word method	21, 43
Progressive alphabet system	3
Random text	108
Reconstruction of equivalent primary components	4, 5, 27, 52, 53
Reconstruction skeletons Footnote 1	26, 56
Relative frequencies	40
Repeating-key ciphers:	
Primary components are different mixed sequences	80
Repeating-key system	3
Analysis of	10
Solution of subsequent messages enciphered by the same primary components	78ff, 80ff
Repetitions:	
Accidental	12, 116
Causal	12
Phenomena of	108
Secondary alphabets	4
Sequence reconstruction skeleton	26
Square tables	5
Symmetry of position	8, 9, 52
Tetragraphic coincidence	117
Theory of factoring	15, 86
Trigraphic coincidence	117
Trigraphs, data pertaining to	117
Types of cipher squares	96ff
Vigenère	9, 19