~~Confidential~~

Register No. 1

WAR DEPARTMENT
OFFICE OF THE CHIEF SIGNAL OFFICER
WASHINGTON

THE INDEX OF COINCIDENCE AND ITS APPLICATIONS IN CRYPTANALYSIS

TECHNICAL PAPER

By

WILLIAM F. FRIEDMAN
Cryptanalyst, Chief of Signal Intelligence Section
War Plans and Training Division

UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1935

THE INDEX OF COINCIDENCE AND ITS APPLICATIONS IN CRYPTANALYSIS [1]

INTRODUCTION

Frequency tables in the analysis and solution of ciphers have commonly been employed to make assumptions of plain-text equivalents for the cipher letters constituting a message. The significance of the various phases of the curves themselves, i.e., the crests and troughs and their relative positions in such frequency tables, has been recognized to some extent, but largely only in connection with the determination of two more or less preliminary points in their analysis: (1) whether the frequency distribution approximates that of a substitution cipher involving only one alphabet or more than one alphabet; (2) whether this approximation corresponds to that of a standard alphabet, direct or reversed, or that of a mixed alphabet.

It will be shown in this paper that the frequency tables of certain types of ciphers have definite characteristics of a mathematical or rather statistical nature, approaching more or less closely those of ordinary statistical curves. These characteristics may be used in the solution of such ciphers to the exclusion of any analysis of the frequencies of individual letters constituting the tables or curves, and *without any assumptions whatever of plain-text values for the cipher letters.*

It is true that cipher systems admitting of such treatment are not very commonly encountered. But inasmuch as such systems are always of a complex nature, which the ordinary methods of cryptanalysis would find rather baffling, a description of a purely mathematical analysis that may be applied to other cases similar to the ones herein described may be considered valuable. In fact, it is possible that the principles to be set forth may find considerably wider application in other phases of cryptanalysis than is apparent at this time.

Two examples of such a treatment will be given in detail: One dealing with a substitution cipher wherein a series of messages employing as many as 125 random mixed secondary alphabets can be solved without assuming a plain-text value for a single cipher letter; the other a multiple alphabet, combined substitution-transposition cipher, solved from a single message of fair length.

[1] The present paper was prepared in 1923. It is a revision of an earlier paper with the same title, published in 1922 by the Riverbank Laboratories, Geneva, Ill. The author takes this opportunity to thank Col. George Fabyan, of the Riverbank Laboratories, for his courtesy in granting permission to publish this revision. Although better methods have been elaborated since the revision was prepared, it has been deemed of interest historically to publish this paper in its 1923 form without change.

PART I

THE VOGEL QUINTUPLE DISK CIPHER [1]

This cipher system involves the use of five superimposed disks bearing dissimilar random mixed alphabets. These disks are mounted upon a circular base plate, the periphery of which is divided into 26 segments; one of these is marked "Plain", indicating the segment in line with which the successive letters of the plain text as found on the five revolving disks are to be brought for encipherment. The remaining 25 segments of the base plate bear the numbers from 1 to 25 in a mixed sequence, which we have called the "Numerical key." This key may, however, consist of less than 25 numbers, in which case one or more of the segments of the base plate will remain blank. The numbers constituting the key are written on the base plate in a clockwise direction beginning immediately at the right of the plain segment (fig. 1).

METHOD OF ENCIPHERMENT

In the accompanying example illustrating the details of encipherment it will be seen that the numerical key consists of 21 numbers, leaving blank, therefore, the four segments immediately preceding the plain segment.[2] Assuming a series of messages, let us suppose the first three begin as follows:

1. Prepare for bombardment at Harvey . . .
2. Enemy attack on Hunterstown . . .
3. Second Field Artillery Brigade . . .

Revolving the five cipher disks successively, and thus bringing the first 5 letters of message 1, PREPA, in line with the plain segment, reading from the outer disk inward in the order 1-2-3-4-5, the cipher letters for this first set of 5 plain-text letters are then taken in the same order from the segments of the disks directly in line with that segment of the base plate that bears the number 1. In this case it is the eighteenth segment after the plain, in a clockwise direction, and, as shown in figure 1, the equivalent cipher letters for this group are MEKJR. The second set of five plain-text letters of message 1, REFOR, are then in a similar manner set in line under the plain segment, and their equivalent cipher letters are taken from the segment immediately following segment 1 of the numerical key, in a clockwise direction, viz, segment 13. The cipher letters in this case are VZQWH. The third group of letters in message 1 finds its cipher equivalents at segment

[1] While on duty in the Code and Cipher Section of the Intelligence Division of the General Staff, G.H.Q., A.E.F., Lt. Col. F. Moorman, Chief of Section, turned over to the writer for study a cipher system together with a series of 26 test messages submitted by Mr. E. J. Vogel, Chief Clerk, who had taken considerable interest in cryptography and had, as a result of his studies, devised the system presented for examination. The writer worked upon the cipher during his leisure moments, but the problem involved considerable labor and solution was not completed before being relieved from duty at that station. The main principles for solution, however, were established and only the detailed work remained to be completed. After an interval of more than a year, while Director of the Cipher Department of the Riverbank Laboratories, Geneva, Ill., the writer turned his attention once more to this cipher and succeeded in completely solving the problem by carrying out those principles to their logical conclusion.

[2] It is recommended that the reader prepare a duplicate of the set of disks in order that he may more readily follow the various steps in the analysis.

18; the fourth, at segment 8. The fifth group of plain-text letters, however, will take its cipher equivalents from the first segment to the right of the plain segment, inasmuch as the segments immediately following segment 8 are blank. Plain-text letters TATHA, therefore, will be enciphered on segment 9, becoming XONJE. This method is continued in like manner throughout message 1. If message 1 contains more than 21 groups of 5 letters, the twenty-second group will take its cipher equivalents on segment 1 again; the twenty-third on segment 13, and so on.

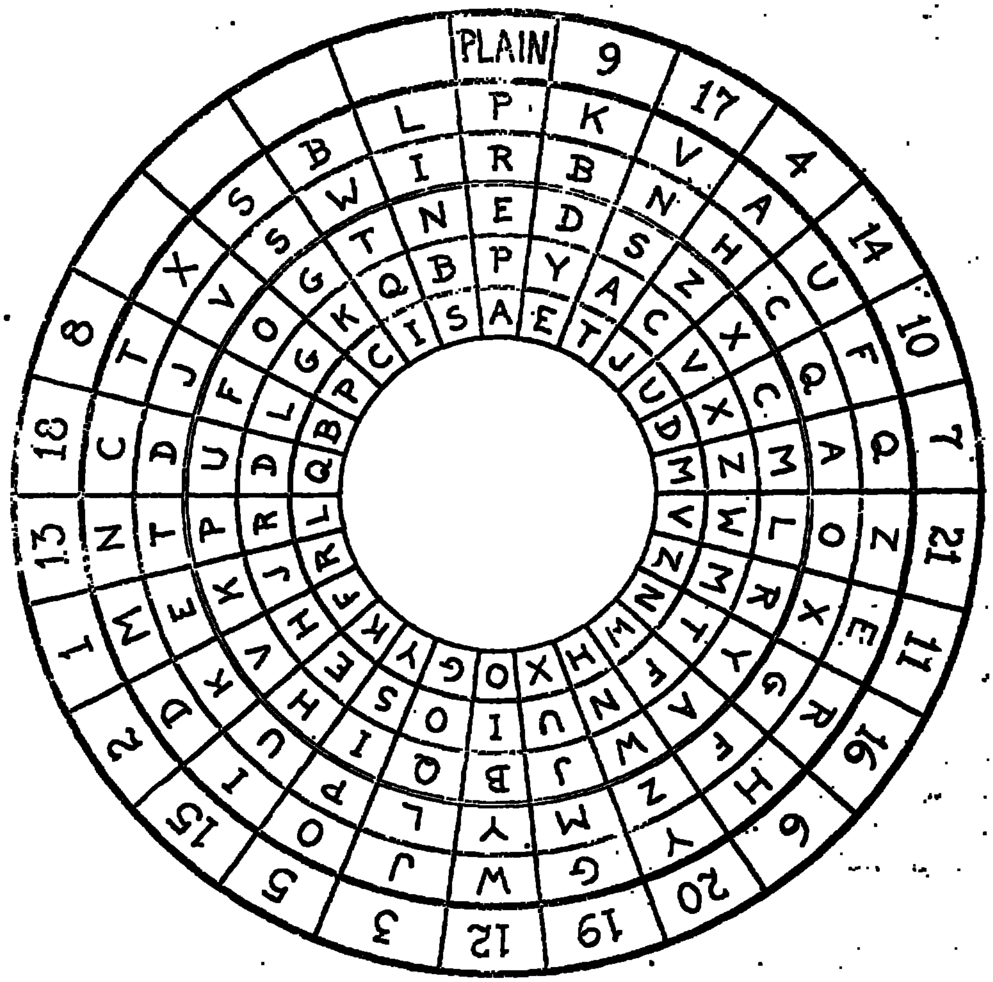


FIGURE 1

Proceeding now to message 2, the first 5 plain-text letters, ENEMY, are set up under the plain segment and the cipher letters are taken from segment 2, becoming LTVBM. The second group of 5 letters is enciphered on segment 1, the third group on segment 13, and so on, throughout the message. The first 5 letters of message 3, SECON, are enciphered on segment 3, becoming YAPAC. The first group of cipher letters in a message is always to be taken from the segment bearing the number which coincides with the serial or accession number of the message in the day's activity.

It will be seen, therefore, that no matter what repetitions occur in the plain-text beginnings of messages, the cipher letters will give no evidences of such repetitions, for each message has a different starting point, determined, as said before, by the serial number of the message in the day's activity. This automatic prevention of initial repetitions in the cipher text is true, however, only of a number of messages equal to the length of the numerical key; for, with a number of messages greater than the length of the key, the initial segments from which the cipher letters are to be taken begin to repeat. In this case, message 22 would necessarily have the same initial point as message 1; message 23, the same as message 2, and so on. Messages 43, 64, 85, . . . , would all begin with segment 1; messages 44, 65, 86, . . . with segment 2, and so on.

The secrecy of the system is dependent upon a frequent change of alphabets and a still more frequent change of numerical key, since it must be assumed, as with all cipher systems or devices, that sooner or later the general method of encipherment will become known to the enemy. The only reliance, therefore, for the safety of the messages must be placed in keeping the specific alphabets and the numerical key *for a given series of messages* from the enemy.

PRINCIPLES OF SOLUTION

The following messages [1] are assumed to have been intercepted within one day and therefore to be in the same alphabets and key:

MESSAGE No. 1

MLVXK QNXVD GIRIE IMNEE FEXVP HPVZR UKSEK MVQCI VXSFW
GVART YBZKJ WVUPV XZCBD BDOLS GHINZ LJCTE KSLPY VPBYD
WTRJK BDDFA ANJXE XGHED ERYVP YPWDJ DFTJV ZHTWB WXTMF
OZDOJ

MESSAGE No. 2

ULJCY GXAEU DTEIL UZBRW GJZSS QLUOX PTFOO NWSHD BPTJO
HQRYY YAXRZ KTEMP UAYMK ISRDZ VUVKW HXAYD YAGSM CURBZ
LBXOV EBBPI BMLCB UMAXF ZSLXV QFXUE MPZMK MQZZT KMURW
EJVB

MESSAGE No. 3

YLMKW CBGSF VGABP HOZFV QQNSQ NQLQL DIGXM XCWAI QFJOQ
TYDEL MBMJB SEPSO DHREM ELKIP KXNMW QYBIH BHFDC GLYWC
YGMMP EEXZH UBBSB SBONG URQKW YRAYU NYUCS LNEMV VNSXN
WGVME MXPDF WGTZE KRLGU ZJFZJ W

MESSAGE No. 4

UFHUJ LTMKJ PONFG RIUGG OZGWS UBNMW WGILB JNXTD BPREX
MMWHB OBFVO TFGSJ SLXEH RTZMI LLUUX FIFWC PGSBA KRCAS
XWKQV SLDKS NTESD QVQBN RZDMB JQLYH LTMXB BSVWL ILKYU
NMFEB HB

MESSAGE No. 5

VJECQ KZJJK DCVQD KSSXY TSUXE GRROH PXKZF ZKFMS VGDWU
XLTBW EXHEF AWQWF ZESMX GWCEM JPNVB GRJGB IBWOH YMOAP
IZYNX GXMSB OZZGK FVURN NJFGQ DTPLV STOID DWVLR TXTBH
WNWIE YJXXW BKOJQ FOUHO T

[1] These are the actual messages submitted by Mr. Vogel.

Message No. 6

XGORF GCHAX DUEIQ XAOWK BKBUH SKWWW XNEYZ JAKUK BEQMG
IWTVU NPCLU KQDIG YLMNC KYJJF SDSTU DCBUK OQUVA WHSXE
VGQSW OGHPI XCYIO SUAEU BQAPY RMDMW FWGSQ HYMRQ LPUVN
LNGQT IPJRI HUMAD DZUTW BFMO

Message No. 7

UFUCL HJYDY ZTHHA NWJWJ IFMZI VZIJE QODUT MZPRX SWNFC
DKXXR TOSVL MNHNO RZRTX QIPFN HOUNL FGUVO TPOWI VHNCQ
RTDCO LNTTM MXMXR TUIIG ZOJCU BTXJK KGUEJ MSJBS ESVWJ
IKGSL APA

Message No. 8

NWWVY UMHVB RPQHF XOHQN IPATI CFMZT DIMQI IRUJI NSBWR
VTEGJ IAFEO BMUUT VPSKO HUYNA VPRXS SCUZB JHCDE WUHJV
GXHRP OIHWF LBTKF QESIO YXVNK AWDAA EQLVJ MYYTH GSEYA
KLXHR NCVNY XSQMC XVXJC TJVSC UJEGZ TFONC HNPM

Message No. 9

NYMPE YZYRZ AWLPP IMPBB VWAXZ QSVFG ITZMT KZNFC DHSUA
NNCMP SOJKI GDPQZ IIPMV ZOCBO UCKXR SEPEM OTZNM AMHWX
NDEUH YUEIP RFOHI QLQHG IJFRT UTNQC JEGAF SQRWO DAJMT
YKXXQ VRQUW

Message No. 10

TEPDA XXHHC FYMFK QRBJV YVJID JBNXF JBLXU KSOXR KNHJK
UFRXL WELQJ QJKFW RLSSF BQJWR BZKYN EAUWP AYSKO UWJDX
CQRVW ZVXXH NZHCW SVEYH NEANW G

Message No. 11

SEYBZ MGSOZ CMPSQ BASFH VFSCG CHKSB ZOPRZ CZEVH OCADC
LGRXO XXCKK MVQGJ XYUOC FFFVJ OZCJE AQSJY WZCMO TOXVM
KEXYA VWONX GTCWT TGLOI IWORT JJVQE HYNK

Message No. 12

OMTXX WUXZE YEHOJ ALJCO EPLPJ RBCVX VVARW DYBDQ FTZDA
XEUJG ROHUP OFSGQ PONLW RAIBP KACIB GMYSB VKHEU XSPFG
WKZTE KZYIZ ZXFJO HZIVG ESGOX YYCEO B

Message No. 13

QNKIT FZHNC GJBSA JIQBX PFTAO RJLUD IKPKI TNUWU EVGHU
LJUFZ UBSMD VNMNZ UWVYQ DJPFY KETMN CLEQS DQXNA GEYHC
JBCPG RNNNH BSYZR STGBV E

Message No. 14

ZJRKK CLYZK RINOK NBTAK NOBBI WRZRX DQTAR AEKMY MOXLT
YVWFL DXZSK ZPWNI JUULI TPUJR TPQGH RJZCQ XCNHU NOKMI
KKGEF FJWWR ZXROX PZUGG HMYNB DUHQZ BJDFA FJAYU RKHKB
ZMKMI D

Message No. 15

OSULC WZAFW SQAUC WPRBI WCFUO JKQHY OIWVX KSMSX MBXOZ
QQDCX CBIMT DGIAS BAQRK RUHHF JYUJG DYNMA CWULT PKWEE
UUHTF FOQOK KNRPI PLLXQ DRDEM JMCFB WLHVX FIFZR STVJV
XDCAS BEHQV OPWEY UTISE NZFCU BXI

Message No. 16

VHXVD RSRYI PVNWQ QEDMJ LTKFN RMDGT DNMBM JDGVP KJMND
RRQDM STTAL GKPPO PZTCU BNCEE HAWMB KRGNU ZXCWQ VEZTC
DIPSG ZUFVH OZIGB TQXND HMHEW VHTLT

Message No. 17

UMMOL HVVEQ GWTJI PQTNS AEFZH TOFQH OOXNQ NPDDT QVSJR
RAABX ZCVWR UUJEG NMJHQ CHZUM TTSIU GHELW HUMFH LZHNH
CJEDW AKHZA SDZIE HIHWG LBUFZ DVPUB DCMXO NAYEY GQHID

Message No. 18

LALNH QUDUA ZBZUD VSJFE MEHXW EUWZT OKWNO OOSIL TASEG
OVQVX PMKKW BQRBI VGDGJ JDAHW RZDIA WQAXB FBRLA AHJEP
PUMEU HJQJP ZSPPQ VZWDL HECDA LPJJS ZOJYB MO

Message No. 19

YTVUL TWEVD MBHKV IHTPI GNXBQ XAUAQ OUFVO GSMKB BAKIG
YRNAF BKIJC ZJSRN WBQHM UYJPT CHCCB RLNVH OLDQA ZZDCV
UWMNZ OPRFC RDONY RCZAM ZYNVQ WFONZ CTTES IRWER GKETA
YUSUK TFECD BMQVB VBWVV VWCZP TWCTJ FHFEH VNDCO MZVLK
YUJPZ BHLDY VVPMD KFHPB VCYU

Message No. 20

OBHOK UWRON AJDFH FRQMI ULOTG XXIEV HMAKV PVMAV OITKD
LDQIW UYVWI JEJRQ MCUZP KGUDN QSPBF TQVZP IZJTU RBUFZ
JUSIZ DCIGX QESJD LIZVM AOPME YNEXI HOXJK KUYHK AORUY
LVD

Message No. 21

OMXOW LTWEW KFJHN ZMSKK GXFDL YLCPT YYYNQ CYYMA ZWRAD
HGWBI HHCGM FDGMN XIQDN NPLYQ NPJNZ OWFVV KMGVH KHCUC
STNVZ CGXHV LZZXX SVTKG POEAC OJYQU MEULH KHYDD ETPDN
QYZVU HIMGG RBHEW CUSO

MESSAGE No. 22

SFDFM	AQMQW	FFONY	GVFZH	JTPSY	IAONR	TWSZJ	OJUGA	HOGXW
YJOUN	AEGQG	HTZPB	YRNEI	BWIVZ	JNPXG	SXXSZ	HVRUC	QEHQR
CXEUF	UKCTB	BETEX	ZSLRK	QFXUV	CLRIL	LHIGG	FVDSQ	RJCSH
JJMIV	JFCEU	QFPMO	TFSOQ	XKESC	CJCFI	BN		

MESSAGE No. 23

LHKGP	KUKNY	WJROY	XZHPK	JJIHU	LUVBU	ZRUDA	XXHRF	MOAFT
XUWVE	YZAJE	NFDEX	GDZGA	JBZET	XYHAL	DGECA	CDPYM	BTMYK
LCSMI	YVSWW	DQNAP	KFPAO	FIQTR	KQQIE	HHYCA	GHEBF	PFTYF
DZAUO	W							

MESSAGE No. 24

OMOOK	OKWHR	MPXQI	PIQMN	ALWNK	HZIKQ	XQNUY	GQZNG	DFTFO
YJODO	XIRIX	KDCBX	UEQTU	VPIWA	NFWOH	GWXXY	EKBMG	MGETD
WFKKZ	PXXZL	XFRWL	FHTEG	INNJI	ETVUP	QHTJY	OP	

MESSAGE No. 25

FPETJ	CDVNY	LMKQU	CDALX	FIYGR	HQMIP	FAONR	QCAVJ	MFIAC
YXDKR	GWMQL	FQMEJ	KOBMS	ZURAN	ULZFE	YDLOT	UZMJM	SETNP
GWILM	FGCVS	NCZGB	HUIRT	XUWAI	DGAML	QBTFK	VYIGT	FLUVK
FJYAZ	YQNVO	WSMSS	CCMFY	KVKRA	Y			

MESSAGE No. 26

XFYKU	NDBFZ	KNDMF	GBVIJ	TKNDA	LGPKS	CWPSK	BSNWH	LTTXN
VDCYO	GYZLI	TRURC	GBIWL	VEKYG	YOOHV	IAUNO	LPSJJ	UTNIE
ZAECE	XQDYB	FDBPB	SEBMA	SRUEU	RUWKE	RIYIT	BDWVG	MZNGB
TRJTG	PZJCM	LSFXW	ANHJO	UZHRQ	ULXFU	XSINC	MVKNT	ZWXXQ
VWRIH	MPMHG	DURHG	IJKIR	UFRNA	APKQP	KAXUF	CACVC	LNICD
ESOWP	MJQEC	GGJBF	SGMFC	TQRGT	JODEE	XHRXE	KIOZG	DLRZV
DLZCS	EMVZL	GAMFQ	ULGXC	WIZWK	IIZYY	HTAVV	OTTTO	RFTHL
HQWVZ	XZUAB	GLHMH	ZFTIQ	OTJEP	VZXCL	YFYOE	LSIJU	SYBMC
YQXGD	SECCE	CIJTI	BUILZ	ZUAPV	QRYXB	VHMWL	LPXGC	RLETI
DJBCW	IHSAM	RHCMJ	RAQBJ	IWORY	OISIE	RESKE	PAUJD	SMQVB
VEASF	TZCWL	AUHKA	MSTZZ	DDQYB	RCMUS	UXWQX	WDPDB	BYGMQ
XRLCH	IOKZO	EPLQU	XJZKI	JNSSI	HCHXX	ZMZKW	PVUGD	QCVQ

It may be of advantage to begin the elucidation of the principles of solution by translating this cipher into terms of the sliding of primary alphabets against one another with the consequent production of a multiplicity of secondary alphabets. For example, by using ordinary sliding alphabets such as are commonly used in cryptanalysis, we may produce the same results as are given by the set of concentric disks. Let us use the alphabets of the illustrative disks, mounted upon sliding strips in pairs, and let us slide each pair of alphabets 8 letters apart. Thus, if we consider the upper one of each pair of alphabets in figure 2 as the plain-text alphabet and begin each alphabet arbitrarily with the letter A, we have the following:

FIGURE 2

1	Plain text	A U F Q Z E R H Y G W J O I D M N C T X S B L $\overline{P}$ K V A U F Q Z E R H Y G W J O I D . . .
	Cipher	A U F Q Z E R H Y G W J O I D $\underline{M}$ N C T X S B L P K V
2	Plain text	A O X G F Z M Y L P U K E T D J V S W I $\overline{R}$ B N H C Q A O X G F Z M Y L P U K E T D . . .
	Cipher	A O X G F Z M Y L P U K $\underline{E}$ T D J V S W I R B N H C Q
3	Plain text	A W J B Q I H V K P U F O G T N $\overline{E}$ D S Z X C M L R Y A W J B Q I H V K P U F O G T . . .
	Cipher	A W J B Q I H V $\underline{K}$ P U F O G T N E D S Z X C M L R Y
4	Plain text	A C V X Z W M T F N U I O S E H J R D L G K Q B $\overline{P}$ Y A C V X Z W M T F N U I O S E . . .
	Cipher	A C V X Z W M T F N U I O S E H $\underline{J}$ R D L G K Q B P Y
5	Plain text	A E T J U D M V Z N W H X O G Y K F R L Q B P C I S $\overline{A}$ E T J U D M V Z N W H X O G . . .
	Cipher	A E T J U D M V Z N W H X O G Y K F $\underline{R}$ L Q B P C I S

Note now that the first set of 5 plain-text letters, PREPA, yields the same set of 5 cipher letters, MEKJR, that we found on page 2 by using the disks. The only thing which these five pairs of independent sliding alphabets have in common in figure 2 is the fact that each pair has been slid apart the same number of letters, viz, 8; if we consider the upper alphabet in each pair as the stationary alphabet, then the lower one has been shifted 8 intervals to the right, or 18 intervals to the left, of the upper alphabet. This corresponds to the position of number 1 in figure 1, for the latter number occupies the eighteenth segment to the right of the plain segment, or the eighth to the left. The relative positions of the numbers in the numerical key, therefore, correspond to the numbers of intervals the primary alphabets in the form of sliding strips would have to be displaced in order to produce the same results as the disks.

Now the sliding against itself of a primary sequence containing 26 letters will give rise to a series of 25 secondary cipher alphabets;[1] likewise, each primary concentric sequence will give rise to a series of 25 secondary alphabets. If the numerical key consists of 25 numbers, all these secondary alphabets will be employed; if it consists of less than 25 numbers, then a correspondingly decreased number of secondaries will be employed.

Since each primary sequence can give rise to a set of 25 secondaries, the total number of possible secondary alphabets in the whole system is 125; but if the numerical key consists of less than 25 numbers, then the total number of secondaries will be less than 125 by exact multiples of 5, since the absence of one or more numbers from the key affects all five primary concentric sequences. For example, if the key consists of 21 numbers, then there will be involved 21×5, or 105 secondary alphabets. In a message of exactly 105 letters, then, each letter will be enciphered by a different secondary alphabet. If the message contains more than 105 letters, then all the letters after the 105th will be enciphered by the same secondary alphabets as at the beginning of the message and in the same sequence.

In the explanation of the method of encipherment it was made clear that the substitution proceeds in a regular manner, taking successive groups of 5 letters; the cipher equivalents are taken from the successive segments, proceeding in a clockwise direction from any given initial segment. It follows, therefore, that in a single long message wherein the complete encipherment requires the passing through of this sequence of segments more than one time, there exist periodic or cyclic phenomena of a type found in various ciphers, due to the presence of a definite or regular *cycle*. In this case, the length of this cycle in terms of groups of 5 letters corresponds exactly with the length of the numerical key; its length in terms of individual letters is five times the length of the key. For the sake of clarity, we shall refer to this cycle when stated in terms of letters as the *period*. Thus, with a key of 21 numbers, the length of the cycle is 21 groups, and the length of the period is 105 letters. If a message consists of 315 letters, for example, the letters would pass through three complete cycles; the 1st, 106th, and 211th letters would be enciphered in exactly similar positions, and therefore by exactly the same secondary alphabet. The 2d, 107th, and 212th letters would likewise be enciphered by the same secondary alphabet, but of course not the same as the preceding secondary alphabet. With a key of 23 numbers, the length of the cycle is 23 groups, the length of the period, 115 letters; the 1st, 116th, and 231st letters would be enciphered by the same secondary alphabet; the 2d, 117th and 232d letters by a different secondary, and so on. If we represent the length of the period by n, then the 1st, $(n+1)$th, $(2n+1)$th, $(3n+1)$th, . . . letters fall in the same secondary alphabet; the 2d, $(n+2)$th, $(2n+2)$th, $(3n+2)$th, . . . letters fall in another secondary alphabet; and so on. If a message be longer than the period, therefore, it will follow that the 1st, 2d, 3d, . . . nth secondary alphabets must contain repetitions of cipher letters, representing

[1] The twenty-sixth secondary alphabet coincides with the normal alphabet, since each plain-text letter would be represented by itself in that secondary alphabet.

repetitions of plain-text letters, for these secondary alphabets, are after all only single mixed cipher alphabets, and the repetition of high-frequency letters in ordinary plain text is a necessary characteristic of all alphabetical languages. Such repetitions will be evidenced by repetitions in the cipher text at n, $2n$, $3n$ intervals, and they may be used to determine the length of the period. Exactly how this is done will presently be demonstrated.

But the determination of the length of the period is only a slight step forward in the analysis. It is true that it will give us the length of the numerical key, but that is all. What we must know next is the sequence of numbers, or rather, the relative positions of the numbers in this key.

We may ascertain this by further scrutiny of the theoretical and actual results of the method of encipherment. It is often the case with various ciphers that the method of encipherment is excellent in principle, and will yield practically indecipherable messages when the messages are very few in number, but the weaknesses in the method are quickly disclosed when it is used for regular traffic such as that necessary in military cryptography, where many messages are to be sent each day in the same key. In the cipher under examination, the weakness is introduced by the fact that the initial segment for each message of the day's activity is determined by the serial number of the message.[1] Now there are as many initial segments for each numerical key as there are numbers in that key. Once the starting point is determined, all the messages pass through the same cycle; different messages merely begin at different points in the cycle. Now, since the numbers applying to these starting points constitute the sequence of numbers in the key, the successive initial segments constitute a series or sequence which, when properly reconstructed, will give us the sequence of numbers in the key.

After the numerical key has been reconstructed, we are yet a long way from solution, for we are still confronted by the more complex problem of reconstructing, or solving, the cipher alphabets.

We have so far analyzed the solution of the problem into the following three steps or phases:

1. The determination of the length of the period.
2. The reconstruction of the numerical key.
3. The reconstruction of the cipher alphabets.

Let us proceed, therefore, to perform each step.

1. **The determination of the length of the period.**—It was explained above how the cipher system will result in the production of repetitions in the cipher text at definite intervals dependent upon the length of the period. The first, $(n+1)$th, $(2n+1)$th, . . . letters fall in the same secondary alphabet; the second, $(n+2)$th, $(2n+2)$th, . . . letters fall in another secondary alphabet; and so on. If there are repetitions in the plain text at n intervals apart, there will be corresponding repetitions in the cipher text. There would be involved here only a slightly modified case of the ordinary process of factoring the intervals between repetitions in the cipher text, as applied in the solution of typical periodic multiple-alphabet ciphers. Thus, in this case, if it happens that the first, second, and third letters of a message, and also the $(n+1)$th, $(n+2)$th, and $(n+3)$th, are the letters THE, then there must be a repetition of the initial trigraph of the cipher text, representing THE, at a distance of n letters. But in a cipher involving so many alphabets as this one, the repetition of trigraphs and polygraphs would naturally be rather infrequent, except in a very long message.

However, the paucity of trigraphs and polygraphs, and even of digraphs, need not prove to be a great obstacle, for the repetitions of individual letters may be used with great accuracy for the same purpose, viz, the determination of the length of the period. The method is based

[1] However, were the initial segments determined in some other manner, the final results would be the same, and the cipher could be solved by a slight modification of method. Even if the initial segments were subject to no law, the cipher could still be solved by the method hereinafter set forth, with some modifications.

upon the construction of what we have called a "Table of coincidence", which will show us mathematically the most probable length of the period. We may as well use the text of our series of messages to illustrate the process.

If we assume the numerical key to consist of 20 numbers, then the length of the period would be 100 letters. Let us write the longest message of our series—viz, message 26—in exactly superimposed lines containing 100 letters each, and then make a count of the recurrences, or more accurately, the coincidences,[1] of letters within the individual columns thus formed.

Note the repetition of the letter F in the second column. This fact is indicated by placing a check mark in the tabulation of coincidences. Where a letter appears three times within the same column (B, in column 8), three check marks are recorded, for there we have a coincidence between the first and second, second and third, and first and third occurrences. Where a letter appears four times in the same column, six check marks are recorded. The number of coincidences for each case corresponds to the number of combinations of two things that can be made from a total of n things.[2]

We note that on the assumption of a period of 100 letters there is a total of 39 coincidences. Now, if the period is really 100 letters in length, then the repetitions of letters within columns are not mere coincidences brought about by chance superimposition of identical letters but are actual recurrences in the restricted sense of being the resultants of the encipherment of similar plain-text letters by the same secondary alphabet. But there is no way of determining from this single tabulation whether the assumption of a period of 100 letters is correct or not, and therefore we do not know whether the repetitions in this case are recurrences or coincidences. This we can determine, however, by a *comparison of tabulations* of coincidences made upon various assumptions of length of period. Theoretically, the correct assumption should yield a higher total of coincidences than the incorrect assumptions, because the recurrence of high-frequency plain-text letters (in English, E, T, O, A, N, I, R, S, H, D) is to be expected, and the number of such causally produced repetitions should certainly be greater than the number of repetitions produced by mere chance in the superimposition.[3]

Let us proceed, therefore, to make a table of coincidence for the various probable lengths of period, first transcribing message 26 into lines corresponding to hypotheses of 105, 110, 115, 120, and 125 letters.[4] Before doing so, however, we find it necessary to introduce a few remarks upon the desirability of using a slight correction factor for this table.

[1] We draw the *distinction* between *recurrences* and *coincidences* on the grounds that the former term should, and will here be used to indicate repetitions of letters in the cipher text causally related to each other by being encipherments of identical plain-text letters by identical alphabets; whereas the latter term indicates repetitions not causally related to each other in this manner but simply the result of chance. A repetition may therefore be either a recurrence or a coincidence. The process of factoring in ordinary multiple-alphabet ciphers of the periodic type has for its purpose the separation and classification of repetitions into the two kinds. Until proved otherwise, all repetitions must be considered coincidences.

[2] The formula is: $C=n(n-1)/2$. Thus, when $n=5$, the number of coincidences is 10; when $n=6$, the number is 15.

[3] Since this paper was written, a further study of the concept of coincidences has made it possible to predict, with a fair degree of accuracy, just how many coincidences should be expected for correct and incorrect assumptions. The mathematical and statistical analyses are given in detail in W. F. Friedman, *Analysis of a Mechanico-Electrical Cryptograph*, Section VI; S. Kullback, *Statistical Methods in Cryptanalysis*, Section VII (Technical Publications, S. I. S., 1934). It results from these mathematical studies that the ratio of the number of actual coincidences to the total number of possible coincidences is .038 for an incorrect case and .066 for a correct one. This knowledge eliminates the necessity for tabulations corresponding to every possible case and gives a reliable means of determining the correct assumption as soon as it is made. (See Notes 1 and 3 on pages 13 and 14 respectively.)

[4] The message need not be written out more than once if long strips of cross-section paper are used, writing a line on each strip. Each line should contain 125 letters, and the various strips can then be arranged to bring the proper letters into superimposition according to each hypothesis in turn.

For really accurate comparison, the totals of coincidence obtained for the various hypotheses should be corrected in order to make proper allowance for the differences in totals due solely to the variation in the number of letters in each column when transcribed according to each hypothesis.[1] From a cryptographic point of view, a total of 100 coincidences in an arrangement where there are 6 letters in each column represents a slightly greater degree of coincidence than in an arrangement of the same message also yielding 100 coincidences, where there are 7 letters in most of the columns; there is less opportunity for coincidences to be produced in the former case. We should, therefore, reduce all the totals of coincidence to some common basis. The reasoning we have followed in the establishment of a correction factor to be applied is as follows:

Message 28 contains exactly 539 letters. When transcribed into lines of 100, 105, . . . , 125 letters, the columns in each of these five set-ups have the following number of letters:

TABLE I

	Period (letters)
39 columns of 6 letters and 61 columns of 5 letters	100
14 columns of 6 letters and 91 columns of 5 letters	105
99 columns of 5 letters and 11 columns of 4 letters	110
79 columns of 5 letters and 36 columns of 4 letters	115
59 columns of 5 letters and 61 columns of 4 letters	120
39 columns of 5 letters and 86 columns of 4 letters	125

Assuming that perfect coincidence can occur in each column (all letters identical), then in a column of 6 letters we can have $6\times5/2=15$ coincidences; in a column of 5 letters, $5\times4/2=10$ coincidences; and in a column of 4 letters, $4\times3/2=6$ coincidences.

If now we find the total number of chances for coincidences for each of the arrangements given in table I, we have the following:

TABLE II

Period (letters)	Conditions	Total chances
100	39 columns of 15 chances each, 61 columns of 10 chances each	1, 195
105	14 columns of 15 chances each, 91 columns of 10 chances each	1, 120
110	99 columns of 10 chances each, 11 columns of 6 chances each	1, 056
115	79 columns of 10 chances each, 36 columns of 6 chances each	1, 006
120	59 columns of 10 chances each, 61 columns of 6 chances each	956
125	39 columns of 10 chances each, 86 columns of 6 chances each	906

Choosing for our basis of comparison the hypothesis of a period of 100 letters, the various proportions of chances for coincidences for each of the remaining hypotheses will constitute correction factors to be applied in each case. They are as follows:

TABLE III

Period (letters)	Chances for coincidence	Correction factor
100	1, 195	1. 00
105	1, 120	1. 07
110	1, 056	1. 13
115	1, 006	1. 19
120	956	1. 25
125	906	1. 32

[1] See footnote 3 on the preceding page. This correction factor is unnecessary if the number of actual coincidences is reduced to a percentage basis, in terms of the total possible number of coincidences.

We are now ready to establish the tables of coincidence for the various hypotheses. Space forbids the actual demonstration of the several arrangements of message 26 to correspond to the various hypothetical key lengths—that shown in figure 3 is typical of them all. We shall give only the final result in table IV.

TABLE IV[1]

Period (letters)	Coincidences on each hypothesis	Total	Correction factor	Corrected total
100	卌 卌 卌 卌 卌 卌 卌 ////	39	1.00	39.0
105	卌 卌 卌 卌 卌 卌 卌 卌 卌	45	1.07	48.2
110	卌 卌 卌 卌 卌 卌 卌 卌 卌 //	47	1.13	53.1
115	卌 卌 卌 卌 卌 卌 卌 卌 卌 卌 卌 卌	60	1.19	71.4
120	卌 卌 卌 卌 卌 卌 卌 /	36	1.25	45.0
125	卌 卌 卌 卌 卌 卌 ///	33	1.32	43.6

There seems to be no doubt but that the period of 115 letters is correct. The cycle, therefore, consists of 115÷5=23 groups, and the numerical key contains 23 numbers. This means that the two final segments bear no numbers, and are therefore blank segments.

2. **The reconstruction of the numerical key.**—Having ascertained the length of the period, and thus the length of the numerical key, the next step is to reconstruct the sequence of numbers constituting the key. As stated before, this process is made possible in this case by the method of encipherment which is such that all the messages of the day's activity go through exactly the same cycle, but the successive messages begin at different initial points in this cycle, and these points coincide with the relative positions of the numbers making up the sequence of numbers in the key.

We do not know the absolute position of any numbers in the numerical key, but we may proceed first to find their relative positions, regarding the key in the nature of a continuous cycle or chain. Later we may find the absolute positions of the numbers in this cycle, i.e., we shall have reconstructed the numerical key itself.

Now, as stated before, all messages proceed through the same cycle; it is only the initial points for the messages which are different. Hence, if we can determine the relative positions in which messages 1, 2, and 3 should be superimposed in order to make all three messages coincide as regards the portion of the cycle through which they pass simultaneously, we shall thus have determined the relative positions of the numbers 1, 2, and 3 in the cycle. For example, if we should find that the first group of message 2 belongs under the twelfth group of message 1, and the first group of message 3 under the sixth group of message 2, we would conclude that the relative positions of these numbers in the cycle are these:

```
    1 2 3 4 5 6 7 8 9 10 11 1 2 3 4 5
  1 . . . . . . . . . . . . 2 . . . . 3
```

[1] As a verification of Note 3, page 11 above, the percentages of the actual number of coincidences to the total possible number has been calculated:

Period	Percentage
100	0.033
105	.040
110	.044
115	.060
120	.038
125	.036

How are these relative positions for messages 1, 2, and 3 to be determined? Clearly, we may use the same basis for this determination as for that concerned with the length of the period--viz, a table of coincidences. For, when messages 1 and 2 are correctly superimposed we should get a higher degree of coincidence between the letters of the superimposed columns than when they are not correctly superimposed. The reasons are the same as in the preceding case: The successive groups of cipher letters in one message represent encipherments by the same secondary alphabets as apply in the other message; hence, repetitions of plain-text letters within columns will result in repetitions of cipher letters within those columns. We can therefore determine the correct superimposition by experiment and the recording of coincidences.

Having found that the key contains 23 numbers, it is obvious that message 24 has the same starting point in the cycle as message 1; we may proceed at once to combine them by direct superimposition. We do likewise with messages 2 and 25, and 3 and 26. The purpose of this step is merely to afford greater accuracy through the increased number of letters with which we shall have to deal in finding the correct relative superimposition of these three sets of messages.

Since message 26 contains the greatest amount of text, we may regard it[1] as our base and try to find the relative position of messages 1 and 24 with respect to it. We now place messages 1 and 24 beneath message 26, beginning the first group of the former beneath the second group of the latter. A tabulation of the number of coincidences in each column is then made. Messages 1 and 24 are again placed beneath message 26, beginning the first group of the former beneath the third group of the latter and again the total number of coincidences is ascertained. In other words, messages 1 and 24 are moved successively 1, 2, 3 . . . 22 intervals[2] to the right of message 26, and a table of coincidences is constructed. The greatest total number of coincidences, as shown in table V, is given when messages 1 and 24 are placed three intervals to the right of message 26. This means, then, that the numbers 1 and 3 occupy these relative positions:

1 2 3
3 · · 1

TABLE V[3]

Intervals	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Coincidences	59	49	86	40	53	39	55	60	59	47	55	45	61	64	53	54	46	58	50	48	52	40

Showing various totals of coincidence when messages 26 and 1 and 24 are superimposed at different intervals corresponding to the successive hypotheses of relative position.

Messages 2 and 25 are next taken for experiment and a similar table of coincidences is made for the various superimpositions with message 26, omitting, of course, the three-interval trial since the position corresponding to that test we found to be occupied by the number 1. As soon as the relative position of each number is found, the subsequent trials corresponding to those numbers may be omitted. The labor is, of course, somewhat tedious but may be done by

[1] It was thought unnecessary to use both messages 3 and 26 as a base, since the latter alone seemed sufficiently long to give reliable information.

[2] An interval in this case is equal to a group of 5 letters, because the encipherment proceeds in segments of 5 letters each. We need try only the first 22 intervals since the cycle consists of only 23 positions, and messages 1 and 24 cannot have the same beginning point as messages 3 and 26.

[3] With regard to the mathematical notion discussed in Note 3 on p. 11, the following remark is pertinent: The total number of coincidences possible when messages 1 and 24 are placed three intervals to the right of message 26 is 1253. The number of actual coincidences, 86, when divided by 1253 gives .068. Since the expected result for a correct assumption is .066, it is at once evident that this assumption is correct and it is consequently unnecessary to consider any further cases.

clerks. This process is continued in similar manner for all the remaining messages. The data for messages 2 and 25 show that they belong five intervals to the right of messages 1 and 24, and the relative positions of the numbers 1, 2, and 3 in the cycle are therefore these:

	1	2	3	1	2	3	4	5
3	·	·	1	·	·	·	·	2

The data for this determination of position for all messages are given in table VI.

TABLE VI.—*Data for determination of the position of the numbers in cycle*

MESSAGE 26 USED AS A BASE

Message																							
1	Position	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	Number of coincidences	59	49	86	40	53	30	55	60	59	47	55	45	61	64	53	54	46	58	50	48	52	40
2	Position	2	3	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
	Number of coincidences	45	40	58	57	62	61	90															
4	Position	2	3	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22	23		
	Number of coincidences	34	35	38	36	16	33	27	23	19	24	28	24	24	21	19	25	20	42	35	32		
5	Position	2	3	5	6	7	8	10	11	12	13	14	15	16	17	18	19	20	22	23			
	Number of coincidences	44	31	33	31	33	29	26	22	16	23	49											
6	Position	2	3	5	6	7	8	10	11	12	13	15	16	17	18	19	20	22	23				
	Number of coincidences	25	33	37	25	40	31	20	11	66													
7	Position	2	3	5	6	7	8	10	11	13	15	16	17	18	19	20	22	23					
	Number of coincidences	23	33	26	27	25	20	24	31	31	16	19	28	28	28	46							
8	Position	2	3	5	6	7	8	10	11	13	15	16	17	18	19	22	23						
	Number of coincidences	36	27	53																			
9	Position	2	3	6	7	8	10	11	13	15	16	17	18	19	22	23							
	Number of coincidences	27	32	26	17	21	21	30	17	13	21	27	58										
10	Position	2	3	6	7	8	10	11	13	15	16	17	19	22	23								
	Number of coincidences	18	18	15	13	23	19	14	18	13	20	10	15	18	52								
11	Position	2	3	6	7	8	10	11	13	15	16	17	19	22									
	Number of coincidences	16	28	10	15	14	16	15	26	17	28	20	31	19									
11[1]	Position		3						13		16		19										
	Number of coincidences		21						30		22		16										
12	Position	2	3	6	7	8	10	11	15	16	17	19	22										
	Number of coincidences	18	27	56																			
13	Position	2	3	7	8	10	11	15	16	17	19	22											
	Number of coincidences	20	19	15	18	18	19	16	38														
14	Position	2	3	7	8	10	11	15	17	19	22												
	Number of coincidences	29	17	32	30	18	20	23	33	24	42												
15	Position	2	3	7	8	10	11	15	17	19													
	Number of coincidences	38	32	37	27	24	59																

[1] Secondary test, using messages 1 and 24 plus 2 and 25 as the base.

TABLE VI.—*Data for determination of the position of the numbers in cycle*—Continued

MESSAGE 26 USED AS A BASE

16	Position	2	3	7	8	10	15	17	19
	Number of coincidences	24	23	47					
17	Position	2	3	8	10	15	17	19	
	Number of coincidences	15	44	30					
18	Position	2	8	10	15	17	19		
	Number of coincidences	48	31	26					
19	Position	8	10	15	17	19			
	Number of coincidences	28	37	06					
20	Position	8	10	17	19				
	Number of coincidences	55	14						
21	Position	10	17	19					
	Number of coincidences	26	33	47					
22	Position	10	17						
	Number of coincidences	48	22						
23	Position	17							
	Number of coincidences	50							

When in any trial the total of coincidences for a certain position stands out prominently from the preceding ones, subsequent trials for the message concerned are omitted.

The final result of carrying out this work for all the messages tried against message 26 is that the following cycle 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
3	18	17	1	8	12	16	20	2	22	15	6	11	5	19	13	23	9	21	7	4	14	10

This reconstructed cycle represents, as stated before, the relative, not the absolute, positions of the numbers in the key, because there is as yet no indication as to what number occupies any given segment on the base disk. Furthermore, we must remember that there is a break of two intervals somewhere within this cycle, representing the two blank segments on the base disk. This fact may cause some difficulty later on but we shall find a way of overcoming it. In the meantime, we may content ourselves with the cycle as established and proceed to an analysis of the results of its reconstruction.

The immediate result is to enable us to superimpose all the messages of the day's activity, as shown in figure 5. We may begin with message 1, or with any other message, but for the sake of convenience in analysis, we may as well transcribe them in regular order.

The letters of each of these 115 columns belong to a corresponding number of secondary alphabets, all different, but all single, mixed, substitution alphabets. Individual frequency tables are made, therefore, and are shown in table VII. The tables are given in groups of five, labeled A, B, C, D, and E, corresponding to the five primary alphabets of the system. The groups of alphabets are given in their proper cyclic sequence, so that each set of five alphabets is accompanied by a number which identifies its position in the sequence of segments. Thus, we may refer to any of these secondary alphabets by number and letter, as for example, 5B, meaning the second single alphabet under segment 5. The A alphabets all apply to the outermost primary alphabet, or alphabet 1; the B alphabets to primary alphabet 2, and so on. We are now ready to attempt an analysis of the cipher text with the object of solving these secondary alphabets and reconstructing the primaries.

The first thought that comes to one is that these individual mixed alphabets may be solved upon the basis of frequency alone, as is commonly done with such frequency distributions. For example, we might assume the most frequently occurring letter in each alphabet to be the equivalent of plain-text letter E, the next, of T, and so on; then substitute the assumed values in the text and try to build up words. But each of these alphabets contains only an average of 36 letters, so that hardly any assumption would carry a considerable degree of certainty. This is especially the case in English text where the letter E does not always stand out prominently as the most frequently used letter in small amounts of text. Were an analysis of this kind absolutely necessary to solution, it is doubtful whether this particular set of messages could be solved except after a long period of patient labor. But it will be shown now that such an analysis is in fact not essential, because we may be able to effect a direct reconstruction of the five primary alphabets, which will not only lead to the solution of all these messages, but will also give us every one of the possible 125 secondary alphabets of the entire system.

TABLE VII

	1				
	A	B	C	D	E
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	B	C	D	E
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					/

TABLE VII—Continued

	12						16				
	A	B	C	D	E		A	B	C	D	E
A	/				/	A	//		/		
B	///	/			𝍸 ///	B	///	//			/
C	//	/	𝍸	///		C	/				
D	//	///		///	/	D	/	/			//
E	/	//		///	///	E	/	///		///	////
F				/	//	F	//		//		///
G	𝍸	/		////		G		//	///		/
H	//	𝍸 /	𝍸			H	/	//		//	////
I		///		//	//	I	///	////	//		
J	/	/	///		/	J	/				
K		//	/	/		K	///				//
L	//		𝍸		/	L	/	/		///	/
M	///	////		//	///	M	///	///		𝍸 ///	//
N	/			/		N			//		///
O	//				//	O		𝍸	//	//	
P	//	///	𝍸			P	/	/	////		
Q		////		///		Q			////		/
R		/	///		/	R	//			/	/
S			/		//	S		//	/	//	
T			//	///		T	//		𝍸	/	𝍸
U	///	/			/	U	/	////			
V				///		V	////	/		////	/
W	///			/	//	W	///	/	///		//
X	/	/	//	𝍸 /	//	X		/	////	////	/
Y	/	/			///	Y		//	/		
Z	/	//	///		/	Z			/		/

TABLE VII—Continued

20						8					
	A	B	C	D	E		A	B	C	D	E
A	//		/	////	///	A	/		//		//
B	/	//		/	/	B					/
C	/	/	/	/		C		///		𝍸	
D	/		//	/		D	𝍸		/		
E		///	///			E			//		/
F	/	/	//	/		F	/	𝍸	/		
G	/	/		/	/	G		//	/		
H			𝍸////		///	H	///		/	/	
I		/	/		𝍸/	I	///		/	//	/
J			///	//	/	J		/	///	/	///
K	𝍸	/	//		////	K	/		//	/	//
L		////	//		//	L	/	𝍸	//		///
M	/	/				M				/	
N		𝍸	/	///	//	N	/		/	////	//
O	𝍸/	///	/	///	𝍸	O	//	////	/	//	///
P				////	//	P	//	////			
Q				////		Q	////	/			𝍸///
R		//	///	/		R	//	/	///	/	///
S		𝍸	/		/	S			/	//	/
T	/	/	/			T	/	/	/	////	
U		//			/	U	𝍸			///	
V		/	/	//	///	V	//	/	/		//
W				/		W		//		/	
X	𝍸/		𝍸/	//		X	/	/	𝍸/	𝍸	///
Y	////		////	///		Y			////		/
Z	////	/	//	/		Z	//	/	//	///	

TABLE VII—Continued

22						15					
	A	B	C	D	E		A	B	C	D	E
A	/	/	//			A	//	/	/		
B		//			/	B	//	///			//
C	////	/	//		/	C			///	/	𝍸
D	/	//	𝍸/	/		D	//				
E	///	////		𝍸////		E			/		
F		//		𝍸		F	/		/	𝍸	
G	////				/	G	///	/		////	///
H		/		//	//	H	/	/			
I		/		/	/	I		/	///	/	𝍸////
J		/	/		///	J	///	//		//	
K	/	/			////	K			///	/	//
L		///	///		////	L	////		/	///	//
M	/	/	////	//	////	M	///	//	𝍸	///	/
N	///		/	𝍸	/	N				////	/
O					/	O	𝍸/	//		/	
P		𝍸/	///	/		P		//		//	
Q	///	/		//		Q	//	////	////	///	/
R		/	/			R	//	/	//		/
S	////		//	/		S		///	/		
T	//	/	/		///	T	/	𝍸			//
U	///	//		/	////	U		/	///	/	/
V	//		///			V	/	///		///	/
W	//	/	/	//		W		/			//
X	/	//	//			X	///		/		//
Y		/	//	//	///	Y		//		/	//
Z	/		/	//	//	Z		/	𝍸//	/	

TABLE VII—Continued

6						11					
	A	B	C	D	E		A	B	C	D	E
A	/	/	////			A			//	///	
B			/	/		B	///	/	///	////	
C	/	/			///	C		////	//	/	////
D	/	//	////			D	////	/		//	
E		/	/			E	/	/	/	//	/
F	𝍸	//		///	//	F	𝍸		////		
G		//	//	/	/	G	𝍸		/	///	
H			//		//	H	/	//	//		////
I	//				////	I		////	//		//
J	/			///		J	/	𝍸 ////	/		
K	//		//	/	/	K	/				/
L	///	𝍸 //		/		L					//
M	/	𝍸 //		//		M		/	/		///
N		/	/	////	/	N	///				
O			𝍸		///	O			//		/
P	/	/	/			P		//			
Q					/	Q		///	/	/	////
R	𝍸 //	//		𝍸 //	/	R	//		//	/	𝍸 /
S	/		////	////		S	//	//	/	////	/
T	//		/	//	///	T	///				/
U	///		𝍸 ///	/		U				///	
V	//	𝍸 /		////		V		///	////		/
W	//	/		/	𝍸 //	W	/	//		𝍸 ///	/
X	//	/	/	/	𝍸	X	/	/	/	//	//
Y					//	Y	//	/	////		
Z		//				Z	//		///	///	//

TABLE VII—Continued

	5						19				
	A	B	C	D	E		A	B	C	D	E
A			/	//	//	A		𝍸	//	////	
B	/	/		///		B				/	
C				///		C	//	//	𝍸	/	/
D	///	/	//	/		D		///	/		///
E	/	/	////	/		E	//	/		//	
F	/					F	/	𝍸	///		
G		///	//	///	//	G		/		/	//
H	////	//			/	H	/	/		𝍸 //	
I	//	//		𝍸 //	/	I	//			/	/
J	///	////	/	/	///	J	//	/	//	//	/
K			///	////		K	𝍸 /				//
L		//		/		L	/				/
M	/	//	///		/	M	/	//	////	//	
N			𝍸		/	N	/		//	𝍸	/
O				//	/	O		/	𝍸///	/	𝍸
P	/	/	//		𝍸	P	///	//	//	/	/
Q	/	///			𝍸 /	Q					//
R	/		///	/		R	///	/			//
S	///		///	/		S	/	/		//	///
T		𝍸 //				T	/	𝍸			
U	//	///	/			U			𝍸	//	//
V	𝍸 /		/	////	𝍸	V		///	/	//	𝍸
W	/	/			/	W	////			//	//
X	/				//	X	////	//	/		/
Y	//	///		///	//	Y	//				
Z	//	/	𝍸 /		///	Z					/

TABLE VII—Continued

13						23					
	A	B	C	D	E		A	B	C	D	E
A	/	///	/		/	A			/	//	///
B	////		//	//		B	///	/		/	/
C	//	////	/			C				/	///
D	///		/	/	~~////~~	D		//	/	//	
E		//	/	/		E				///	/
F	/	/		//		F	/	~~////~~	//		/
G			/		//	G			/	///	//
H	/	/	//	////	///	H		//	//	/	/
I	/	//		///	/	I		/	//	////	/
J	/				///	J		///	//	//	/
K		//	///	//		K	////	//	///	////	
L	///	/		/		L	///	/		/	//
M		/	/	///	///	M	///		/		
N	//	////		/		N				/	
O	//	//			////	O	////		//	/	//
P			//		/	P		~~////~~/			/
Q	~~////~~/		~~////~~/	~~////~~		Q		//	//		///
R	////	/	/		//	R	/	/		////	//
S	/		////		/	S	////	/	~~////~~//		//
T	//	//	////	////	///	T	///	//	~~////~~		/
U	/	/	//	////	/	U	/		/		/
V			///	//	//	V	///	///		/	/
W		~~////~~				W	/	/	////	//	//
X	//			/	/	X		//		///	//
Y		//			//	Y					///
Z		///	//	/		Z		/			

TABLE VII—Continued

9					
	A	B	C	D	E
A				//	𝍸/
B		/	/		/
C	//			/	////
D			/		//
E			𝍸/		𝍸/
F	//	//			
G	𝍸/		/		/
H	𝍸	𝍸			/
I	/	/	///	////	//
J		//			
K	//	///	//	/	
L		/		///	
M			///		
N	///	//	///	////	
O	/	///		/	/
P	//		/	///	
Q		///		/	
R			////		//
S		//	/	𝍸/	
T	//	////	/	/	
U	𝍸	////	/		
V			/	//	
W	/		////	///	////
X	//	/		//	/
Y	//	//	///	//	//
Z					///

21					
	A	B	C	D	E
A	//	////		//	
B		/	//	𝍸	/
C	/			////	
D	/				
E		///	//	/	///
F			/	//	//
G	𝍸	/	/		
H	/	//			///
I		/	/		/
J	////	𝍸/	///		
K			//	/	//
L	/				/
M	/	/	///	/	
N	/	/			//
O	//	//	/	𝍸	
P	/	//	/		////
Q	/		//	///	//
R	//	/	////	////	
S			//		/
T			/	//	
U		//	//	///	////
V	/	//			
W	/	//			///
X	/	/	𝍸//	////	//
Y	𝍸/		//		//
Z	𝍸	𝍸			////

TABLE VII—Continued

	7						4				
	A	B	C	D	E		A	B	C	D	E
A	//	////		//		A		////	//		
B	卌 //		/		卌	B	/	//	////	/	卌 ////
C	/	//	//	////		C	///	/	//		//
D			///		//	D	//	/		卌 /	////
E	//	////	///	////	/	E			///	/	/
F	/	卌			///	F	//	///	//		
G	/		/	/	////	G	/	///			
H			////		卌	H	卌	//	/	///	//
I			/	/		I	//		//	/	
J				///	/	J	///	卌	//	/	//
K	////	//	/		//	K	卌	/			/
L	//	/	//	/	/	L			//		
M	//		///	卌		M	/	卌		卌 /	
N	/	//		/		N				/	//
O			/	/	//	O	//	//			///
P	卌 /			///	///	P		/	/	//	/
Q	/	////	/	/	//	Q		/	///		/
R	/		/			R	/				
S	//	/				S	//			//	//
T	/	///	//			T		/	//	/	//
U	/	/	卌 //	/	/	U	///	/	///	///	//
V				/		V	/	/		卌	
W		/	/	/	//	W		/	/		/
X	///	//			/	X					
Y				//	/	Y			卌	///	/
Z		卌 /		////		Z	//	/	/		

TABLE VII—Continued

	14						10				
	A	B	C	D	E		A	B	C	D	E
A	/		///	//	///	A		//			////
B		//	/	///	///	B	𝍸 /			/	
C			/	/		C	/		////	/	
D	//		/	/	/	D		//	///	////	
E	/	///		///		E	//	/		/	//
F		/			//	F			/	𝍸 /	
G	/	///	/	/	/	G	///				𝍸
H			𝍸	/		H			/		
I	//		//		//	I			/	///	
J		/	//	//	/	J		/	/		//
K	/	////	/	𝍸 /	///	K	/		/	///	///
L	𝍸			/		L		///			𝍸 /
M	/	/	///	/		M					
N	//	/	/	///	//	N	/	/	/		
O	/		/		/	O		/	/		
P		/	////	/		P	//		//		////
Q	/			/	//	Q	//	//	///		/
R		//	𝍸			R		////		//	
S	//	𝍸	/	/	//	S	//	𝍸			
T	/	𝍸 /				T	////		/		
U		////			𝍸	U	//	/	///	/	///
V	𝍸		/	/	//	V	𝍸		𝍸	//	
W	////	/	/	/	/	W		///	/	𝍸 //	////
X			/	////		X	//	𝍸	//		
Y	//					Y		/	/		
Z	𝍸			/	////	Z	//	///	///	////	/

TABLE VII—Continued

3						18					
	A	B	C	D	E		A	B	C	D	E
A	//		///	/	//	A	//	//	//	/	
B						B	/	卌/	/	/	//
C			//		////	C	/			/	///
D		/		/	/	D	/	/			/
E	//	//		//	////	E				卌	/
F	/	////			卌	F	///		//	卌//	卌
G			///	///	////	G	/	/	///		///
H	////		卌//	///		H		//	///		//
I	卌	卌			/	I		//	///	//	
J	/	//	//		//	J	/	///	///	卌/	
K		//	/	////	/	K	///	//		///	//
L		////	/		//	L	卌/		/		
M	//		////	/		M	//	/	//		//
N		//	//	/		N	////		/	//	
O	/			///		O	/	/		/	
P			/	/		P		//	/		
Q			//			Q			/		/
R	////	//	/	/		R	////				
S				/		S	/	//		////	///
T	/	/		////	///	T			////		//
U		///	/		卌/	U	/		卌/		/
V	/	//				V	/				
W	/	/	/	///	/	W		/		//	/
X	////	////		//		X	//	/	/		
Y	卌//		//	////		Y	//	卌/	///		
Z		/	///	/		Z		卌		/	卌//

17						17					
	A	B	C	D	E		A	B	C	D	E
A			/		////	N	/	////	///	/	/
B	/	///	///	///	///	O		///		//	///
C	///			/	///	P			/		///
D	////	///	////		/	Q	////		/	/	
E	/	///	/		/	R		//	//		
F	/	/	///	/	/	S	////		卌////		
G		//			/	T			/	/	/
H	//	/			/	U	///	卌/		卌	/
I	/			//	///	V	//	//		卌	//
J	/	/	/	///	/	W			//		/
K	/		/			X	/	/			/
L	/				/	Y		///	//	//	
M		//	//	卌////		Z	卌/				//

The method which we are about to demonstrate is based upon the fact that the segments from which the cipher groups are taken follow one another from a given initial point in a regular succession, uninterrupted in this case except for a break of three segments representing the two blank segments of the key plus one blank which is always present representing the plain segment. To explain the principle of this method in detail, attention is directed to the fact that, as a result of the system of encipherment, the series of successive cipher equivalents for any given plain-text letter in any one of the five primary alphabets coincides with the sequence of letters in that alphabet. The series will coincide with the complete alphabet except for the omission of 1, 2, 3, or more letters depending upon the number of blank segments. For example, turn to figure 1 and note that, in alphabet 1, the sequence of letters beginning with A is as follows: A U F Q Z E R H Y G W J . . . .

Now it is patent that if we place letter A *of the first primary alphabet in the plain segment, its series of successive cipher equivalents coincides with the sequence of letters succeeding* A *in the same alphabet, viz,* U F Q Z E R H Y G W J O I D M N C . . . .

If we place another letter of the same primary alphabet—*for* example, Z—in the plain segment, its series of successive cipher equivalents constitutes exactly the same sequence, except with a different initial point, viz, E R H Y G W J O I D M N C . . . . In other words the successive cipher equivalents for these 2 plain-text letters come from one and the same cycle or sequence. Now, the same is true with respect to every other letter of alphabet 1, and also of the other primary alphabets. Of course, the sequence is different for each primary alphabet.

Since this cycle or sequence of letters is the same for all the letters of each primary alphabet, only the series of successive cipher equivalents for *one* letter of each primary alphabet is necessary in order to effect a complete reconstruction of that alphabet. In other words, if we can select with accuracy the cipher equivalent for *one and only one* plain-text letter in each of the successive 115 secondary alphabets, we can then arrange these equivalents into five sequences of letters which will coincide with the five primary alphabets, thus resulting in their reconstruction. The reconstructed sequences will be complete except for the omission of one or more letters representing the blank segments. If the numerical key consists of 23 numbers, three letters will be missing from each sequence. These letters will be known, of course, but their relative positions in the omitted section will have to be found later.

Obviously, the letter which will lend itself best to such a procedure is E, for it is the most frequently occurring letter in English text. If, therefore, by a careful study of the individual frequency tables applying to the columns of the superimposed messages, we can select the cipher equivalent of only the letter E with certainty in the successive *secondary* alphabets, we shall at once have the sequences of letters in the five primary alphabets and the solution of the problem will be at hand. For example, if in a hypothetical sequence of these alphabets we select the letters K, N, Q, and V, respectively, as the four successive cipher equivalents of E, then this will mean that in primary alphabet 1 there is a sequence . . . K N Q V . . ., providing a break in the numerical key does not exist between the members of the sequence of key numbers applying to the segments concerned. Continuing this process, ultimately the five primary alphabets can be completely reconstructed. But we must remember always that this process is dependent upon the correct assumptions for the cipher equivalent of E in each of the 115 secondary alphabets, or columns of cipher text.

Let us attempt such a reconstruction. Turning to the series of secondary alphabets given in table VII, we try to find in each alphabet the letter which undoubtedly represents plain-text letter E. At the very start we encounter difficulties. In alphabet 1A, the letters M and Y are of equal frequency. There is no way of telling which letter represents E, so that we shall have to consider both M and Y as possibilities. In alphabet 8A again we have difficulties, for both J and Q have the same frequency. It begins to look like a very doubtful procedure. As we go further along, the difficulties in selecting the representative of E increase rather than decrease and the cryptanalyst becomes lost in a multiplicity of possibilities. Evidently this method, as the preceding one, while theoretically correct, is practically out of the question because of the limited size of each frequency table. In fact, it is doubtful whether we can select the representative of E with certainty in any one of the A alphabets,[1] and certainly, if we cannot do this with the letter that theoretically occurs the most frequently, we cannot do it with any other letter.

It was at this point, when apparently a blank wall confronted the writer, and there seemed little hope of solution, that he evolved the method which finally resulted in solution, and which embodied such new principles that he was led to describe them in this paper. This method had recourse to some simple mathematics, easy of comprehension and application when the underlying principles have been grasped.

First, let us make what we have termed a "consolidated frequency" table for all of the secondary alphabets applying to the first, or A, primary alphabet. This is done by collecting the data contained in the individual frequency tables shown in table VII into one large table, taking only the data applying to the letters of primary alphabet 1. This larger table is shown below (table VIII).

[1] It was found later that the cipher equivalent of E has the greatest frequency in only 3 out of the 23 alphabets. In one alphabet E did not occur at all, and in six cases it occurred only two times. It will be of interest to the reader to study these tables for the information they contain with regard to the extreme degrees of variation from the normal that small frequency tables can exhibit.

TABLE VIII.—*Consolidated frequency table for alphabet 1*

Cipher letter	Segment 1	8	12	16	20	[illegible]	22	15	6	11	5	19	[illegible]	23	9	21	7	4	14	10	3	18	17	Total frequency	Number of segments occupied	Average frequency per segment
A		2	1	2	3	1	1	2	1				1			2	2		1		2	2		22	14	1.58
B			3	4	[illegible]			2		3	1		4	3			7	1		6		1	1	37	13	2.85
C	1	1	2	1	[illegible]		4		1			2	2		2	1	1	3		1		1	3	27	16	1.69
D	4	2	2	1	[illegible]	6	1	2	1	4	3		3			1		2	2			1	4	39	17	2.30
E	1	2	1	1			3			1	1	2					2		1	2	2		1	20	13	1.54
F		1		2	[illegible]	[illegible]		1	5	5	1	1	1	1	2		1	2			2	3	1	30	17	1.76
G	3	1	5		[illegible]		4	3		5					6	5	1	1	1	3		1		40	14	2.86
H	3	1	2	1		[illegible]		1		1	4	1	1		5	1		5			4		2	35	15	2.34
I				3		[illegible]			2		2	2	1	5	1			2	2		5		1	29	12	2.42
J		4	1	1				3	1	1	3	2	1			4		3			1	1	1	27	14	1.93
K				3	5	[illegible]	1		2	1		6		4	2		4	5	1	1		3	1	40	15	2.67
L	1		2	1		[illegible]		4	3			1	3	3		1	2		5			6	1	34	14	2.43
M	5		3	3	[illegible]		1	3	1		1	1		3		1	2	1	1		2	2		31	16	1.94
N	1	3	1			1	3			3		1	2		3	1	1		2	1		4	1	28	15	1.86
O	3	2	2		6	2		6					2	4	1	2		2	1		1	1		35	14	2.50
P		1	2	1		2			1		1	3			2	1	6			2				22	11	2.00
Q	1	4				4	3	2			1		6			1	1		1	2			4	30	12	2.50
R	1			2	[illegible]	2		2	7	2	1	3	4	1		2	1	1			4	4		38	16	2.37
S		3					4		1	2	3	1	1	4			2	2	2	2		1	4	32	14	2.28
T	1	3		2	[illegible]	1	2	1	2	3		1	2	3	2		1		1	4	1			31	17	1.82
U		3	3	1		5	3		3		2		1	1	5		1	3		2		1	3	37	15	2.47
V	2	1		4		2	2	1	2		6			3		1		1	5	5	1	1	2	39	16	2.45
W	3	1	3	3			2		2	1	1	4		1	1	1			4		1			28	14	2.00
X			1		6	1	1	3	2	1	1	4	2		2	1	3			3	4	2	1	38	17	2.24
Y	5	1	1		4					2	2	2			2	6			2		7	2		36	12	3.00
Z	2		1		4	2	1			2	2					5		2	5	2			6	34	12	2.83
																								841		

Average frequency per cipher letter $= \frac{841}{26} = 32.4$ occurrences.

This consolidated frequency table is of a rather peculiar nature. Each column gives the frequency of the cipher letters in a particular segment and there are 23 such columns, corresponding to the 23 segments of the numerical key. The numbers of the columns are determined by, and coincide with, the sequence of numbers in the cycle as given on page 16, viz, 1, 8, 12, 16, etc. Each row gives the frequency of a particular cipher letter in the successive segments and since the columns succeed one another in the cyclic sequence, it follows that the frequencies in the successive segments on a line with any given cipher letter form a definite *sequence of frequencies*. There being 26 cipher letters, there are 26 such rows or sequences of frequencies. The total frequency for each cipher letter is given in the column labeled as such and the average frequency for all cipher letters is then found to be 841÷26=32.4 occurrences. The number of different segments in which the cipher letter applying to any given line occurs is indicated in the next column; and the average frequency per segment for each cipher letter is given in the last column.

Before we can proceed it will be advisable to establish certain principles which will enable us to follow the subsequent reasoning more easily. We shall make use of alphabet 1 shown in table VIII, calling attention to the fact that the same principles apply to the other four primary alphabets. In order to make the illustration comparable in all its details with the real situation in the test problem, let us make the numerical key 23 numbers in length by adding numbers 22 and 23 at the end of the key shown in figure 1.

Let us see what successive plain-text letters the cipher letters A, B, and C represent in the sequence of segments.

FIGURE 6.—*Plain text*

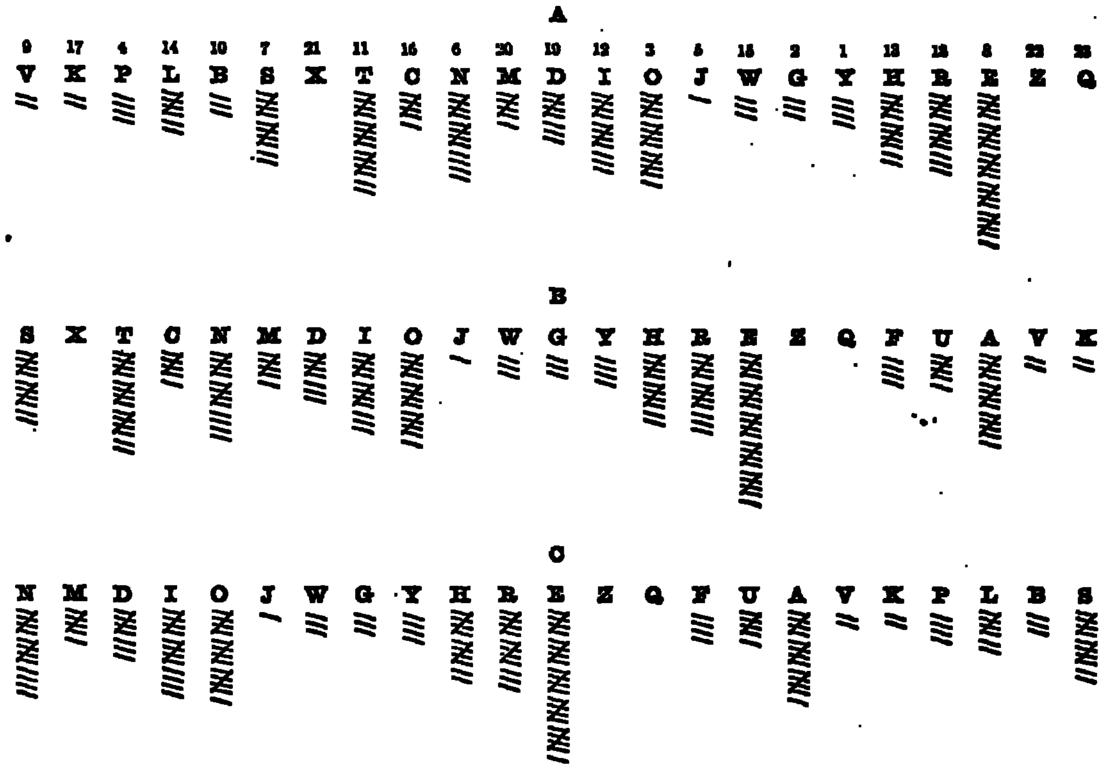
A—V K P L B S X T C N M D I O J W G Y H R E Z Q
B—S X T C N M D I O J W G Y H R E Z Q F U A V K
C—N M D I O J W G Y H R E Z Q F U A V K P L B S

It will be noted that the successive plain-text letters which cipher letters A, B, and C represent constitute almost exactly the same sequence in the three lines. This follows from the nature of the cipher system itself, and the cause of it has already been pointed out. In the B line there is a section not present in the A line, consisting of the letters FUA; in the A line, the section not present in the C line consists of the letters XTC; and in the C line, the section not present in the B line consists of the letters PLB. This is due to the interruption in the numerical key; the section omitted will consist of 3 sequent letters in each case, but these letters will be different for every cipher letter.

Let us now accompany the sequence of the plain-text letters opposite each of the letters A, B, and C, with a sequence of frequencies corresponding to their normal theoretical frequencies [1] for English text.

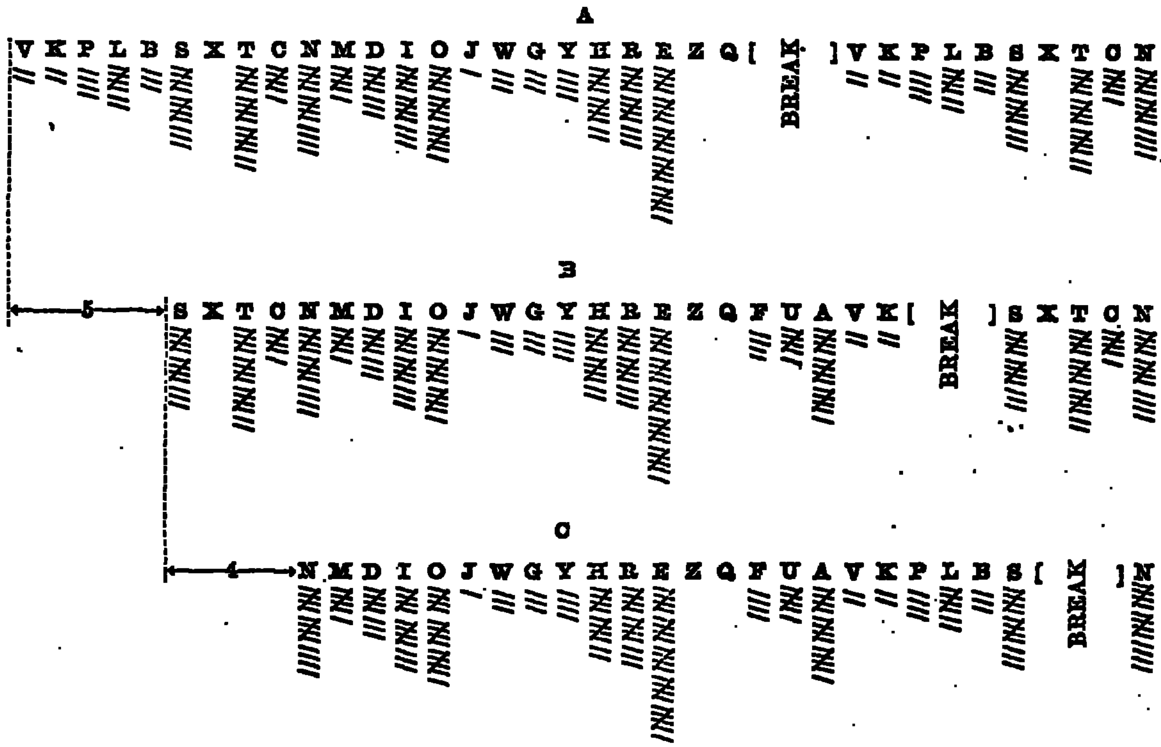
[1] These theoretical frequencies are given by Hitt on the basis of 200 letters of plain text. See Hitt, Parker. *Manual for the Solution of Military Ciphers*, 1918, p. 6.

FIGURE 7



Now, since the sequences of plain-text letters represented by these sequences of frequencies are the same, it follows that we can so arrange the latter as to make the successive individual frequencies coincide; and if we make due allowance for the break in the sequences caused by the omitted sections of 3 letters, the three sequences should coincide exactly. Thus:

FIGURE 8



In order to make the sequences coincide, we displaced the B sequence five intervals to the right of the A sequence, and the C sequence four intervals to the right of the B sequence. Let us reverse the order of these letters, A, B, and C, and space them in accordance with the number of intervals which each sequence of frequencies has been shifted relative to the others. Thus:

4 3 2 1 6 5 4 3 2 1
C . . . B . . . . A

Refer now to the illustrative cipher alphabet in figure 1 and note that this corresponds to the order of these letters A, B, and C in this primary alphabet. We have determined the order of these letters in our alphabet merely by correctly superimposing or shifting the three sequences of frequencies relative to one another so as to make the individual frequencies coincide.

Now had we not known what letters these individual frequencies in each sequence of frequencies represented but had merely been given the sequence of frequencies themselves, it would still have been just as easy to find the correct relative positions of the three sequences from a comparison of the positions of high and low points in each *sequence of frequencies*. In other words, *we do not need to know what letters the individual frequencies in each sequence of frequencies represent*; it is still possible to determine the relative positions (in the primary alphabet) of the letters applying to each sequence in the cipher alphabet *by a study of the positions of the high and low points in each sequence of frequencies*. No analysis whatever of the individual frequencies is necessary, the entire frequency table being treated as an ordinary statistical curve. This, in its final analysis, is the meaning of the proposition stated in the opening paragraph of this paper.[1] It thus follows that the five alphabets of our problem may be reconstructed, without a knowledge of what letter any individual frequency in the sequences of frequencies (as shown in table VIII) represents, by an analysis of these frequency tables considered as true statistical curves.

Let us return now to the test messages. Table VIII represents a set of 26 sequences of frequencies similar in origin to those for A, B, and C in the illustrative alphabet above. We could superimpose these sequences in the same manner and as easily as we did in the messages themselves were it not for two circumstances: First, we know that there is an interruption of three blanks in the cycle which we have reconstructed but do not know where these blanks must be inserted. Consequently, some allowance must be made for the blank segments in each sequence of frequencies. Secondly, the individual frequencies in each sequence of frequencies in our problem do not exactly correspond to the theoretical frequencies of the plain-text letters to which they apply but only correspond approximately to the theoretical. In some cases this approximation is far from close because of the paucity of text, and this will make the determination of the correct relative positions of two sequences a much more difficult process than was the case with the illustrative sequences above.

We are, therefore, confronted with the problem of superimposing the sequences of frequencies correctly without a knowledge of these two factors, and this we shall accomplish by a slight modification of method and a recourse to some simple mathematics.

First, as to the modification of method due to our ignorance of the exact location of the break of three intervals in the numerical key: this consists in superimposing sequences, not to find the relative positions of *any pair of sequences* but to find such sequences as are *one and only one interval apart;* i.e., sequences which represent a relative displacement of only one interval. The reason for this step is now to be explained.

Let us consider the sequence of theoretical frequencies corresponding to the cipher letter A and the letter which immediately follows it in the illustrative alphabet, viz, U, arranging the two sequences as though we had only reconstructed the cycle and had not as yet determined the numerical key. Let us begin both sequences with segment 9, the first segment in the key.

---

[1] The ordinary frequency table applying to a plain text or a cipher alphabet does not correspond to the ordinary frequency distribution of statistical work. In the latter, the position of the points along one of the axes of the graph and their extension along the other axis are either causally related, or the curve treats of data which, being subject to the operation of the laws of probability, form the normal, or Quetelet, curve of error. In the former, the positions and extensions of the coordinates are not related in any way unless one considers the arbitrary order of the letters of the alphabet as constituting a cause. The positions of the coordinates in a cryptographic curve were determined many centuries ago when the English language was first evolved.

But the sequences of frequencies in table VIII are not similar in origin to the ordinary plain-text or cipher alphabet frequency tables of cryptographic work. They are, in fact, closely related to certain frequency distributions of statistical data because the position and extensions of the coordinates are absolutely determined by a cause other than the arbitrary order of the letters of the English alphabet. These two characteristics of the curves of a series of secondary alphabets may be varied at will by changing the sequence of letters in the primary alphabet. Any set of frequency distributions applying to a series of secondary alphabets derived from a variable primary alphabet may be treated in the same mathematical manner as these will be treated in the subsequent pages.

FIGURE 9

A

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

U

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

It is evident that we may superimpose these sequences correctly by shifting the A sequence one space to the right of the U sequence or the U sequence one space to the left of the A sequence. Thus:

FIGURE 10

A

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

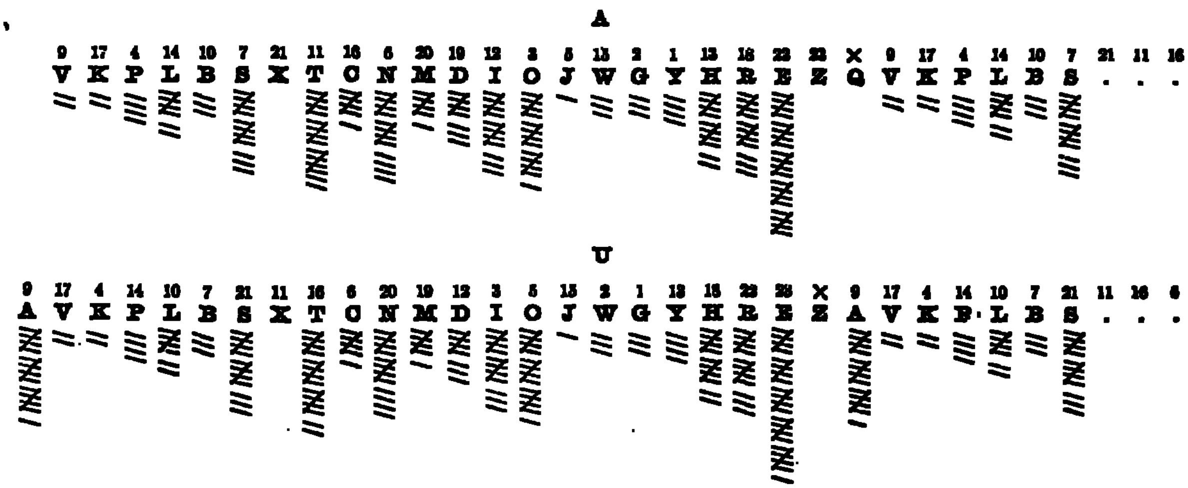
U

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

Note now that complete and perfect coincidence between successive pairs of superimposed segments is obtained except in two cases, viz, those involving the letters A and Q in the U and A lines, respectively. This is due to the fact that the break in the sequence comes between segments 23 and 9. The frequency of letter A of the U sequence should be matched with the frequency of A in the A sequence, but the latter does not occur because of the break in the numerical key. The same is the case with the letter Q of the A sequence.

Now suppose that we did not know where the break in the numerical key falls, and let us superimpose the sequences again. Thus:

FIGURE 11



It is seen that perfect coincidence is still maintained throughout except in the case of the one pair of segments containing the letters A and Q of the U and A lines, respectively. By omitting the three blank segments representing the place where the break occurs, we have brought the letters A and Q into an incorrect superimposition. But the amount of error due to the superimposition of one pair of incorrect segments as against the correct superimposition of 22 other pairs is of so little consequence that it may be neglected altogether. In other words, when we superimpose sequences *which are only one interval apart*, relative to each other, we may neglect the discrepancy that would be due to our ignorance of where the break in the numerical key comes.

Now suppose that we did not know that the letter U immediately follows A in this illustrative primary alphabet, and had only a table containing the frequencies applying to the cipher letters (similar to those shown in table VIII). It is evident that by placing the A sequence one interval to the right of all other sequences successively, and choosing that sequence which most closely coincides with the A sequence in the positions of the high and low points, we shall thus have determined what letter immediately follows A in the primary alphabet; this follows because the letter applying to that sequence of frequencies occupies a definite position in the cipher alphabet, viz, it follows A. In this case, the U sequence would be chosen and we would conclude that the sequence in the primary alphabet is . . . A U . . . . Taking the U sequence, the same operation is performed with the other sequences as before, and we thus find the letter that follows U in the primary alphabet. Theoretically, therefore, we should be able to reconstruct the complete primary alphabet by this method, and thus overcome the difficulty due to our failure to know exactly where the interruption in the numerical key falls.

In this process of matching sequences of frequencies to decide which one most closely coincides with a given sequence, we cannot depend upon a mere ocular examination and comparison. We must reduce the operation to a mathematical method. This, we shall proceed to consider.

Let us return to table VIII and select that line for experiment which gives the best indications of representing the closest approximation to a theoretical frequency table containing as few elements as are contained in the average line in the table. In a theoretical frequency table of small size such as the one shown in figure 12 only the high-frequency letters are represented; the low-frequency letters are absent. The average frequency per letter that does occur is evidently the total frequency, viz, 34, divided by the number of different letters that go to make up this total, viz, 12. This quotient is 34÷12=2.83.

FIGURE 12

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
////		/	/	~~////~~ /			//	///			/		///	////			//	//	~~////~~						

Total frequency=34
Number of different letters=12
Average frequency $=\frac{34}{12}=2.83$

Note now that the Y line of frequencies in table VIII averages 3.00 occurrences per segment; that is, the average frequency per plain-text letter which Y represents is 3.00. This is even a little better than the average theoretical frequency per letter as determined above. Let us consider the Y sequence of frequencies, therefore, as representing the closest approximation to a theoretical frequency table of a similar total of occurrences.

Following the method discussed above, let us see if we can find the letter in alphabet 1 which follows Y.

Taking our Y sequence of frequencies, let us apply it to all the other sequences, placing the Y sequence one interval to the right of the other sequences. Thus, with A the Y sequences are placed in these relative positions:

FIGURE 13

		1	8	12	16	20	3	22	15	6	11	5	19	13	23	9	21	7	4	14	10	3	18	17	1	
A			2	1	2	2	1	1	2	1				1			2	2		1		2	2			
	Y		5	1	1		4					2	2	2			2	6			2		7	2		
			1	8	12	16	20	2	23	15	6	11	5	19	13	23	9	21	7	4	14	10	3	18	17	1

We shall now proceed to find an abstract number such as will indicate the degree to which these two sequences of frequencies agree, or fit, when placed with reference to each other as in figure 13. It is evident that when we strike the letter which really follows Y in alphabet 1, the corresponding sequence of frequences concerned should give the best fit with the Y sequence and thus produce the greatest degree of coincidence.

Let us now compare the two superimposed sequences above, segment by segment. In the upper one of the first pair of superimposed segments there are 2 occurrences of A; in the lower one, 5 occurrences of Y. The first pair of segments agree, therefore, in 2 occurrences; i.e., there are 2 coincidences. In the next pair of segments, an occurrence of 1 in the Y sequence is matched by an occurrence of 1 in the A sequence; i.e., there is 1 coincidence. Let us go through the rest of the segments in the same manner. The results are given in figure 14.

FIGURE 14

			1	8	12	16	20	2	22	15	6	11	5	10	13	23	9	21	7	4	14	10	3	18	17	1	
	A			2	1	2	2	1	1	2	1				1			2	2		1		2	2			
		Y		5	1	1		4					2	2	2			2	6			2		7	2		
				1	8	12	16	20	2	22	15	6	11	5	10	13	23	9	21	7	4	14	10	3	18	17	1
Coincidences				2	1	1	0	1	0	0	0		0	0	1			2	2		0	0	0	2	0		

We have a total of 12 coincidences. But we must also take into consideration the number of noncoincidences between the two sequences; for these, as can easily be demonstrated, are of equal importance with the coincidences. In the two hypothetical sequences given below, both with the same total frequency, the number of coincidences is very high, viz, 28, yet the two sequences do not agree closely at all, for there are 45 noncoincidences.

FIGURE 15

	5		2		6		4		5	1	2	6		2		4		5		4		5
	5			5			4		5		2	6		2	6		4		2	4	5	
Occurrences (101)	10	0	2	5	6	0	8	0	10	1	4	12	0	4	6	4	4	5	2	8	5	5
Coincidences (28)	5						4		5		2	6		2						4		
Noncoincidences (45)	0		2	5	6		0		0	1	0	0		0	6	4	4	5	2	0	5	5

Let us find the total of noncoincidences, therefore, between the Y and the A sequences. In the first pair of segments there are 3 noncoincidences; in the second pair, none; in the third pair 1, etc. Let us now add these to our table, and include also the number of occurrences in the segments, for we shall have need of this information very soon.

FIGURE 16

			1	8	12	16	20	2	22	15	6	11	5	10	13	23	9	21	7	4	14	10	3	18	17	1
	A			2	1	2	2	1	1	2	1				1			2	2		1		2	2		
		Y		5	1	1		4					2	2	2			2	6			2		7	2	
				1	8	12	16	20	2	22	15	6	11	5	10	13	23	9	21	7	4	14	10	3	18	17
Coincidences				2	1	1	0	1	0	0	0	0	0	0	1	0	0	2	2	0	0	0	0	2	0	=12
Noncoincidences				3	0	1	2	3	1	2	1	0	2	2	1	0	0	0	4	0	1	2	2	5	2	=34
Occurrences				7	2	3	2	5	1	2	1	0	2	2	3	0	0	4	8	0	1	2	2	9	2	=58

We find that the total of coincidences is 12, that of noncoincidences, 34. The difference is $12-34=-22$. Were the sequences in closer agreement, this difference would be a positive quantity; but as a rule, we shall find it to be a negative quantity in our work because of the fact that the frequencies are relatively low throughout. In this case, then, the number of noncoincidences is 22 greater than the number of coincidences. This difference between the totals of coincidences and noncoincidences will be used as the basis for the determination of the degree to which two sequences coincide, and inasmuch as we shall have a great many such differences to compute, a short cut to their determination will be of use. If we subtract the total of occurrences from three times the total of coincidences, we can find this difference directly without having to count up the number of noncoincidences. Thus, in this case, $(3\times 12)-58=-22$. In all subsequent determinations we shall use this method.

Now it is obvious that the number of coincidences as well as the number of noncoincidences is not only a function of the distribution of the occurrences in each sequence of frequencies but also of the total number of occurrences. It is patent that in one pair of sequences with a greater total number of occurrences than in another pair, the totals of coincidences and noncoincidences might be greater in the former than in the latter from the mere fact that there are more opportunities for coincidence and noncoincidence in the former case. We should therefore take into consideration the total number of occurrences in the two superimposed sequences, and the most logical correction would be to divide the difference between the totals of coincidences and noncoincidences by the total number of occurrences of all the segments. For example, it is only reasonable to place more reliance upon a case in which out of 30 occurrences the difference between the totals of coincidences and noncoincidences is +10, than upon a case in which out of 60 occurrences the difference between these same totals is also +10. In the former case, the quotient obtained by dividing the difference, +10, by the total occurrences, 30, is +.33; in the latter case, the quotient obtained by dividing +10 by 60 is only +.17, only half as much.

To this quotient, which indicates in a general way the "goodness of fit" of the two superimposed sequences, and which is obtained by dividing the difference between the totals of coincidences and noncoincidences by the total occurrences, we have applied the name "Index of coincidence".[1] It is evident that the greater the index of coincidence, the better is the agreement between the superimposed sequences, and thus, the closer is the fit. Where the two sequences are relatively low in frequency, the total of noncoincidences will, as a rule, be greater than the total of coincidences, so that the difference will usually be a negative quantity and the index will also be negative. As these negative indices approach 0, they become closer to positive indices, so that when we are dealing with negative indices, the lowest absolute index will indicate the greatest coincidence. Thus, an index of −.03 will indicate a much better fit than an index of −.35.

Returning now to the case in hand, we found the difference between the totals of coincidences and noncoincidences to be −22. Since a total of 58 occurrences enters into the formation of these two tables, then the index of coincidence for the assumption that A follows Y in alphabet 1 is $-22\div 58=-.38$.

Let us perform the same calculations for the rest of the letters in table VIII (p. 30), omitting, of course, Y. The data are given in table IX.

[1] See S. Kullback, loc. cit., Sections VI and VII for other and more reliable tests for matching alphabets.

TABLE IX

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	58	12	−22	−.38	J	63	15	−18	−.29	S	68	14	−26	−.38
B	73	15	−28	−.37	K	76	13	−37	−.40	T	67	11	−34	−.51
C	63	13	−24	−.38	L	70	16	−22	−.28	U	74	18	−20	−.27
D	75	16	−27	−.36	M	67	9	−40	−.60	V	75	12	−39	−.52
E	56	12	−20	−.36	N	64	16	−16	−.24	W	64	7	−43	−.67
F	66	11	−33	−.50	O	71	10	−41	−.58	X	74	16	−26	−.35
G	76	8	−52	−.69	P	58	17	− 7	−.12	Z	70	11	−37	−.53
H	71	13	−32	−.45	Q	66	17	−15	−.23					
I	65	10	−35	−.54	R	74	15	−29	−.34					

As stated above, the best fit is obtained when the index of coincidence is the greatest positive quantity. In none of the cases above is the index of coincidence positive, but the value for P, viz, —.12, approaches the nearest to a positive quantity, and therefore represents the greatest degree of coincidence. The next greatest index is given by the letter Q; but inasmuch as the index for P is almost twice as great as that for Q, we may conclude that it is the letter P, and not the letter Q, which immediately succeeds Y in the alphabet 1.

We may now proceed to find the letter that follows P. The same operations are performed with the letter P as with the letter Y[1], this time using the frequency of P as the base and trying it one interval removed from the frequencies of all letters except Y and P, for, as the position of each letter is determined, it can be automatically omitted from the succeeding calculations. The data are as follows:

TABLE X

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	44	11	−11	−.25	I	51	8	−27	−.53	R	60	14	−18	−.30
B	50	10	−20	−.40	J	49	11	−16	−.36	S	54	8	−30	−.56
C	49	13	−10	−.20	K	62	12	−26	−.42	T	53	11	−20	−.38
D	61	11	−28	−.46	L	56	8	−32	−.57	U	60	9	−33	−.55
E	42	9	−15	−.36	M	53	11	−20	−.38	V	61	7	−40	−.65
F	52	10	−22	−.42	N	50	9	−23	−.46	W	50	8	−26	−.52
G	62	10	−35	−.57	O	57	9	−30	−.53	X	60	11	−27	−.45
H	57	13	−18	−.32	Q	52	7	−31	−.60	Z	56	8	−32	−.56

It is seen that the index for letter C, viz, —.20, represents the greatest degree of coincidence. But the index for A is —.25 and that for R is —.30. In other words, the index for C is only .05 greater than that for A, and .10 greater than that for R. The question then arises: Is a difference of .05 or .10 in favor of C over A and R, respectively, a significant difference? In

[1] The method which has been given here for determining the index of coincidence is not very effective when the alphabets are very much different in size. The mathematical study mentioned in Note 3 on page 11 has provided a method which is effective in such cases, and with its help a great deal of work could have been saved in the following calculations. For, as soon as the proper position of the P sequence of frequencies with respect to the Y sequence had been determined, the two could have been combined so as to yield a base of twice as many letters. Similarly, the knowledge that P is followed by C would permit an additional adjunction of frequencies to the base. Since the accuracy obtained in matching increases with the number of letters involved in the test, the further results could have been obtained with much less difficulty.

other words, might not the letter A or R follow P, instead of C? The answer may be found by modifying the method in one particular. We have been superimposing sequences with a relative displacement of but one interval. If now we superimpose sequences with a relative displacement of two intervals, the error, due to the failure to take into account the break in the numerical key, will be greater than it is with a relative displacement of one interval, for now there will be two incorrect pairs of superimposed letters, but still the results will be significant. Let us, therefore, test out all the letters which are less than twice the index of C, with the Y sequence removed *two intervals*. Thus with A, the sequences are in this position:

FIGURE 17

A			2	1	2	2	1	1	2	1				1			2	2		1		2	2			2
	Y		5	1	1		4						2	2	2			2	6			2		7	2	

The data for the letters which may possibly follow P, when tested with the Y sequence at two intervals removed, are as follows:

TABLE XI

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	58	9	−31	−.55	H	71	16	−23	−.32	R	74	11	−41	−.55
C	63	20	− 3	−.05	J	63	10	−33	−.52	T	67	13	−28	−.42
E	56	13	−17	−.30	M	67	14	−25	−.37					

These calculations show conclusively that C follows P in alphabet 1. In the tables showing the calculations for the reconstruction of alphabet 1, whenever the first calculation, using one-interval data, fails to show a letter whose index of coincidence is at least twice as great as its nearest rival, a secondary calculation will be made using two-interval data. In two cases, viz, for the letters following K and Z, it will be noted that three-interval data were employed to determine the correct letter subsequent to an inconclusive secondary calculation.

The tables containing the rest of the data for the reconstruction of alphabet 1 are given below:

TABLE XII.—*Data for reconstruction of primary alphabet No. 1*

Y

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	58	12	−22	−.38
B	73	16	−27	−.37
C	63	13	−24	−.38
D	75	16	−27	−.36
E	56	12	−20	−.36
F	66	11	−33	−.50
G	76	8	−52	−.60
H	71	13	−32	−.45
I	65	10	−35	−.54
J	63	15	−18	−.29
K	76	13	−37	−.40
L	70	16	−22	−.28
M	67	9	−40	−.60
N	64	16	−16	−.24
O	71	10	−41	−.58
P	58	17	− 7	−.12
Q	66	17	−15	−.23
R	74	15	−29	−.34
S	68	14	−26	−.38
T	67	11	−34	−.51
U	74	18	−20	−.27
V	75	12	−39	−.52
W	64	7	−43	−.67
X	74	16	−26	−.35
Z	70	11	−37	−.53

P

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	44	11	−11	−.25
B	59	10	−29	−.49
C	49	13	−10	−.20
D	61	11	−28	−.46
E	42	9	−15	−.36
F	52	10	−22	−.42
G	62	19	−35	−.57
H	57	13	−18	−.32
I	51	8	−27	−.53
J	49	11	−16	−.36
K	62	12	−26	−.42
L	56	8	−32	−.57
M	58	11	−20	−.38
N	50	9	−23	−.46
O	57	9	−30	−.53
Q	52	7	−31	−.60
R	60	14	−18	−.30
S	54	8	−30	−.56
T	53	11	−20	−.38
U	60	9	−33	−.55
V	61	7	−40	−.65
W	50	8	−26	−.52
X	60	11	−27	−.45
Z	56	8	−32	−.56

Y at two intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	58	9	−31	−.55
C	63	20	− 3	−.05
E	56	13	−17	−.30
H	71	16	−23	−.32
J	63	10	−33	−.52
M	67	14	−25	−.37
R	74	11	−41	−.55
T	67	13	−28	−.42

C

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	49	14	− 7	−.14
B	64	14	−22	−.34
D	68	18	−12	−.18
E	47	9	−20	−.43
F	57	13	−18	−.32
G	67	15	−22	−.33
H	62	14	−20	−.32
I	56	11	−23	−.41
J	54	13	−15	−.28
K	67	11	−34	−.51
L	61	18	− 7	−.11
M	58	17	− 7	−.12
N	55	12	−19	−.35
O	62	20	+ 2	+.03
Q	57	11	−24	−.42
R	65	16	−17	−.26
S	59	10	−29	−.49
T	58	15	−13	−.22
U	65	9	−38	−.58
V	66	16	−18	−.27
W	55	14	−13	−.24
X	65	13	−26	−.40
Z	61	13	−22	−.36

TABLE XII.—*Data for reconstruction of primary alphabet No. 1*—Continued

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
		O		
A	58	13	−19	−.33
B	72	11	−30	−.54
D	74	17	−23	−.31
E	55	11	−22	−.40
F	65	15	−20	−.31
G	75	13	−36	−.48
H	70	13	−31	−.44
I	64	13	−25	−.39
J	62	9	−35	−.57
K	75	16	−27	−.36
L	69	16	−21	−.30
M	66	12	−30	−.46
N	63	17	−12	−.19
Q	65	14	−23	−.35
R	73	14	−31	−.43
S	67	15	−22	−.33
T	66	17	−15	−.23
U	73	25	+2	+.03
V	74	17	−23	−.31
W	64	14	−22	−.34
X	73	13	−34	−.47
Z	69	9	−42	−.61

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
		U		
A	50	12	−23	−.39
B	74	15	−29	−.39
D	76	15	−31	−.41
E	57	13	−18	−.32
F	67	13	−28	−.42
G	78	25	−3	−.03
H	72	15	−27	−.38
I	66	13	−27	−.41
J	64	15	−19	−.30
K	77	13	−38	−.49
L	71	14	−29	−.41
M	68	19	−11	−.16
N	65	14	−23	−.35
Q	67	9	−40	−.60
R	75	16	−27	−.36
S	69	12	−33	−.48
T	68	15	−23	−.34
V	76	16	−28	−.37
W	65	21	−2	−.03
X	75	14	−33	−.44
Z	71	17	−20	−.28
		O at two intervals		
G	75	26	+3	+.04
W	63	12	−27	−.43

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
		G		
A	62	16	−14	−.23
B	77	16	−29	−.38
D	79	15	−34	−.43
E	60	12	−24	−.40
F	70	13	−31	−.44
H	75	15	−30	−.40
I	69	14	−27	−.39
J	67	19	−10	−.15
K	80	14	−38	−.48
L	74	15	−29	−.39
M	71	16	−23	−.32
N	68	10	−38	−.56
Q	70	12	−34	−.49
R	78	16	−30	−.39
S	72	13	−33	−.46
T	71	13	−32	−.45
V	79	20	−19	−.24
W	68	11	−35	−.52
X	78	17	−27	−.35
Z	74	8	−50	−.68
		U at two intervals		
A	59	17	−8	−.14
J	64	14	−22	−.34
V	76	18	−22	−.29

TABLE XII.—*Data for reconstruction of primary alphabet No. 1*—Continued

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A				
B	59	13	−20	−.34
D	61	15	−16	−.26
E	42	8	−18	−.43
F	52	14	−10	−.19
H	57	11	−24	−.42
I	51	9	−24	−.47
J	49	9	−22	−.45
K	62	17	−11	−.18
L	56	13	−17	−.30
M	53	12	−17	−.32
N	50	9	−23	−.46
Q	52	8	−28	−.54
R	60	18	−21	−.35
S	54	12	−18	−.33
T	53	11	−20	−.38
V	61	18	−22	−.36
W	50	8	−26	−.52
X	60	15	−15	−.25
Z	56	12	−20	−.36
G at two intervals				
D	79	14	−37	−.47
F	70	17	−19	−.27
K	80	20	+7	+.09
X	78	21	−15	−.19

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
K				
B	77	11	−44	−.57
D	79	23	−10	−.13
E	60	8	−36	−.60
F	70	13	−31	−.44
H	75	20	−15	−.20
I	69	12	−33	−.48
J	67	11	−34	−.51
L	74	13	−35	−.47
M	71	9	−53	−.75
N	68	15	−23	−.34
Q	70	20	−10	−.14
R	78	16	−38	−.49
S	72	13	−33	−.46
T	71	14	−29	−.41
V	79	17	−28	−.35
W	68	11	−35	−.51
X	78	16	−30	−.39
Z	74	23	−5	−.07
A at two intervals				
D	61	15	−16	−.26
H	74	11	−41	−.55
Q	69	8	−45	−.65
Z	73	12	−37	−.50
G at three intervals				
D	79	28	+5	+.06
Z	74	18	−20	−.27

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
D				
B	76	15	−31	−.41
E	59	17	−8	−.14
F	69	13	−30	−.44
H	74	16	−26	−.38
I	68	15	−23	−.34
J	66	15	−21	−.32
L	73	16	−25	−.34
M	70	19	−13	−.19
N	67	16	−19	−.28
Q	69	16	−21	−.30
R	77	17	−26	−.34
S	71	22	−5	−.07
T	70	22	−4	−.06
V	78	24	−6	−.08
W	67	20	−7	−.10
X	77	16	−29	−.38
Z	73	14	−31	−.43
K at two intervals				
E	60	13	−21	−.35
S	72	20	−12	−.17
T	71	18	−17	−.24
V	79	28	+5	+.06
W	68	16	−20	−.30

TABLE XII.—*Data for reconstruction of primary alphabet No. 1*—Continued

V

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	76	14	−34	−.45
E	59	16	−11	−.19
F	69	13	−30	−.44
H	74	15	−29	−.39
I	63	11	−35	−.52
J	66	12	−30	−.45
L	73	10	−43	−.59
M	70	13	−31	−.44
N	67	17	−16	−.24
Q	69	12	−33	−.48
R	77	16	−29	−.38
S	71	14	−29	−.41
T	70	20	−10	−.14
W	67	15	−12	−.19
X	77	26	+ 1	+.01
Z	73	13	−34	−.47

X

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	75	15	−30	−.40
E	58	11	−25	−.43
F	68	20	− 8	−.12
H	73	19	−16	−.22
I	67	17	−16	−.24
J	65	15	−20	−.31
L	72	20	−12	−.17
M	69	16	−21	−.30
N	66	15	−21	−.32
Q	68	16	−20	−.29
R	76	28	+ 8	+.11
S	70	15	−25	−.36
T	69	16	−21	−.30
W	66	11	−33	−.50
Z	72	13	−33	−.46

V at two intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
F	69	15	−24	−.35
L	73	19	−16	−.22
R	77	26	+ 1	+.01

R

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	75	16	−27	−.36
E	58	10	−28	−.48
F	68	21	− 3	−.04
H	73	11	−40	−.55
I	67	15	−22	−.33
J	65	10	−25	−.39
L	72	18	−18	−.25
M	69	14	−27	−.39
N	66	18	−12	−.18
Q	68	14	−26	−.38
S	70	23	− 1	−.01
T	69	19	−12	−.17
W	66	10	−36	−.55
Z	72	14	−30	−.42

S

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	69	13	−30	−.44
E	52	12	−16	−.31
F	62	12	−26	−.42
H	67	20	− 7	−.10
I	61	14	−19	−.31
J	59	14	−17	−.29
L	66	13	−27	−.41
M	63	17	−12	−.19
N	60	12	−24	−.40
Q	62	9	−35	−.57
T	63	12	−27	−.43
W	60	16	−12	−.20
Z	66	13	−27	−.41

R at two intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
H	73	23	− 4	−.06
M	69	16	−21	−.30
W	66	13	−27	−.41

H

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	72	9	−45	−.63
E	55	14	−13	−.24
F	65	14	−23	−.35
I	64	10	−34	−.53
J	62	15	−17	−.27
L	69	18	−15	−.22
M	66	15	−21	−.32
N	63	17	−12	−.18
Q	65	12	−29	−.45
T	66	15	−21	−.32
W	63	20	− 3	−.05
Z	60	16	−21	−.30

W

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	65	20	− 5	−.08
E	48	10	−18	−.38
F	58	13	−19	−.33
I	57	7	−36	−.63
J	55	13	−16	−.29
L	62	11	−29	−.47
M	59	12	−23	−.39
N	56	14	−14	−.25
Q	58	14	−16	−.28
T	59	18	− 5	−.09
Z	62	10	−32	−.52

H at two intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B	72	25	+ 3	+.04
T	66	17	−15	−.23

TABLE XII.—*Data for reconstruction of primary alphabet No. 1*—Continued

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B				
E	57	8	−33	−.58
F	67	15	−22	−.33
I	66	23	+3	+.05
J	64	12	−28	−.44
L	71	11	−38	−.54
M	68	15	−23	−.34
N	65	8	−41	−.63
Q	67	5	−52	−.78
T	69	14	−27	−.39
Z	71	11	−38	−.54
I				
E	49	10	−19	−.39
F	59	11	−26	−.44
J	56	6	−38	−.68
L	63	13	−24	−.38
M	60	9	−33	−.55
N	57	20	+3	+.05
Q	59	10	−20	−.40
T	61	15	−16	−.26
Z	63	12	−27	−.43
N				
E	48	11	−15	−.31
F	58	10	−28	−.48
J	55	15	−10	−.18
L	62	13	−23	−.37
M	59	16	−11	−.19
Q	58	14	−16	−.28
T	60	13	−21	−.35
Z	62	14	−20	−.32
I at two intervals				
J	56	13	−17	−.30
M	60	10	−30	−.50
Q	59	17	−8	−.14
Q				
E	50	11	−17	−.34
F	60	9	−33	−.55
J	57	9	−30	−.53
L	64	13	−25	−.39
M	61	20	−1	−.02
T	62	14	−20	−.32
Z	64	6	−46	−.72
M				
E	51	10	−21	−.41
F	61	20	−1	−.02
J	58	14	−16	−.27
L	65	13	−26	−.40
T	63	18	−9	−.14
Z	65	11	−32	−.49
F				
E	50	8	−26	−.52
J	60	13	−21	−.35
L	64	12	−28	−.44
T	62	13	−23	−.37
Z	64	18	+8	+.13
Z				
E	54	16	−6	−.11
J	64	9	−37	−.58
L	68	9	−41	−.60
T	66	16	−8	−.12
F at two intervals				
E	50	14	−8	−.16
T	61	15	−16	−.26
M at three intervals				
E	51	16	−3	−.06
T	62	18	−8	−.13
E				
J	47	13	−8	−.17
L	61	12	−25	−.41
T	59	9	−32	−.54
J				
L	61	18	−7	−.11
T	58	12	−22	−.38
L				
T	65	22	+1	+.02

The now completed alphabet 1, which is written tentatively on a disk and mounted upon the base, 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
Alphabet 1	Y	P	C	O	U	G	A	K	D	V	X	R	S	H	W	B	I	N	Q	M	F	Z	E	J	L	T

This reconstruction is, so far, purely the result of hypothesis. Moreover, we do not as yet know where the break in the numerical key comes, and this we shall proceed now to ascertain. There are several methods. We might, for example, assume that the first segment to the right of the plain segment bears the number 1 of the numerical key. Then, turning to the secondary alphabet applying to that segment in table VIII (the first column), we note what the plain-text equivalents for the letters A, B, C, . . . Z, would be on this hypothesis by actual trial with this alphabet mounted on the base disk, comparing the frequencies given with their normal expectancy. Thus, the letters D, M, and Y—with frequencies of 4, 5, and 5, respectively—would represent the letters K, Q, and T, respectively; this would be far from a good agreement with expectancy. Hence, we would conclude that the first segment of our base disk does not bear the number 1. Let us assume that the second number of our cycle applies to the first segment of our base disk, and proceed again to note the plain-text letters corresponding to this assumption. The cipher letters with their frequencies and equivalents would be as follows:

FIGURE 18

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

The agreement of these frequencies with their normal expectancy is fair, although not striking. There are too many occurrences of low-frequency letters, G, P, K, and Q, and although the frequencies for E, O, I, N, and R are excellent, there are not enough occurrences of the high frequency letters, T, A, and S. We must give this hypothesis further scrutiny. If the first segment bears the number 8, then the second segment would bear the number 12. Again let us match the frequencies given with their corresponding plain-text letters on this hypothesis.

FIGURE 19

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

This assortment of letters is also fair, but still the evidence is not conclusive. We might continue along these lines, attempting to establish definitely that the numerical key begins with the number 8, or does not. But let us try another method.

Referring now to the Y sequence of frequencies in table VIII still retaining the assumption that the numerical key begins with the number 8, the Y sequence should be broken and rearranged in order to conform to the assumed break in numerical key in this manner:

FIGURE 20

	8	12	16	20	2	22	15	6	11	5	19	13	23	9	21	7	4	14	10	3	18	17	1	(Break)			
Y	1	1		4					2	2	2			2	6			2		7	2		5				

Below these frequencies let us place the plain-text letters which they represent upon the assumption that the break does occur between the numbers 1 and 8 in the key.

FIGURE 21

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

This too, looks like a fairly good assortment of high-frequency letters, with the single exception of T. It may be that we have really struck the correct place for the break after all. Let us corroborate it by still another method.

Assuming the break to be as shown in figure 20 let us make a list of the cipher letters which should represent E in the successive segments. Thus:

FIGURE 22

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

Let us turn now to the respective frequencies of these cipher letters as given in table VII, taking the frequencies of the letters J, L, T, . . . in the A columns, successively beginning with segment 8.

FIGURE 23

	8	12	16	20	2	22	15	6	11	5	19	13	23	9	21	7	4	14	10	3	18	17	1	[Break]		
Equivalents of plain-text E	J	L	T	Y	P	C	O	U	G	A	K	D	V	X	R	S	H	W	B	I	N	Q	M			
Frequency	4	2	2	4	2	4	6	3	5		6	3	3	2	2	2	5	4	6	5	4	4	5			

On the whole, this is as good as can be expected for the small number of occurrences in each table. Let us do the same for the letters T, O, and A. If the results are as good as those for E we may conclude that we have really found the absolute positions of the numbers in the numerical key:

FIGURE 24

	8	12	16	20	2	22	15	6	11	5	19	13	23	9	21	7	4	14	10	3	18	17	1	[Break]	
Equivalents of plain-text T	Y	P	C	O	U	G	A	K	D	V	X	R	S	H	W	B	I	N	Q	M	F	Z	E		
Frequency	1	2	1	0	5	4	2	2	4	0	4	4	4	5	1	7	2	2	2	2	3	6	1		
Equivalents of plain-text O	U	G	A	K	D	V	X	R	S	H	W	B	I	N	Q	M	F	Z	E	J	L	T	Y		
Frequency	3	5	2	5	5	2	3	7	2	4	4	4	5	3	1	2	2	5	2	1	6	0	5		
Equivalents of plain-text A	K	D	V	X	R	S	H	W	B	I	N	Q	M	F	Z	E	J	L	T	Y	P	C	O		
Frequency	0	2	4	6	2	4	1	2	3	2	1	6	3	2	5	2	3	5	4	7	0	3	3		

These distributions and frequencies are certainly excellent and we may regard our numerical key as established. The initial segment after the plain is number 8. (See Addendum, p. 56.)

We could proceed to reconstruct alphabets 2, 3, 4, and 5 by exactly the same principles as were used in the reconstruction of alphabet 1. To do so would be purely of theoretical interest because it would represent a case where the reconstruction of five primary alphabets, from the frequencies of 115 unknown secondary alphabets, is accomplished without a preliminary tentative decipherment of even so much as a single word. In short, it would represent a case where the cryptanalyst, without attempting the decipherment of any part of the text, comes at once, after such a reconstruction as the above, to be in the same position as the correspondents, and can decipher any message as rapidly as the legitimate recipients.

Where a staff of clerks and experts is available, this method would indeed be followed, for the personnel could be divided into five groups, each group being assigned an alphabet to reconstruct. Each group could be subdivided into two sections, one working forward from a given letter, the other working backward from the same letter. After not more than 3 to 6 hours all five alphabets will have been reconstructed in their entirety. Or perhaps it would be more practicable to reconstruct only three of the primary alphabets, say the first, third, and fifth, filling in the other two from the resulting decipherment.

In the present instance, however, only alphabets 1 and 3 were reconstructed, the reconstruction of alphabet 3 being successfully accomplished by the application of the same principles as were used for alphabet 1. Then a partial decipherment, in which the repetitions of digraphs and trigraphs within adjacent columns played an important part, led to the reconstruction of the other three primary alphabets.

The data for the reconstruction of alphabet 3 are given in table XIV. Since the location of the break in the numerical key was determined after the reconstruction of alphabet 1, the columns in table XIV, upon which the reconstruction of alphabet 3 is based, are given in the correct numerical key order so that any two sequences of frequencies could be superimposed at any intervals. However, one-interval and two-interval data were used almost exclusively except in one or two doubtful instances where greater intervals were employed.

Table XIII.—*Consolidated frequency table for alphabet 8*

Cipher letter	Segment 8	12	16	20	2	22	15	6	11	5	19	13	23	9	21	7	4	14	10	3	18	17	1	Total frequency	Number of segments occupied	Average frequency per segment
A			1	1	2	2	1	4	2	1	2	1	1				2	3		3	2	1	1	30	17	1.76
B	1							1	3			2		1		1	4	1			1	3	1	19	11	1.74
C	2	5		1		2	3		2		5	1			2	2	2	1	4	2			3	37	15	2.47
D	3			2	1	6		4		1	1	1	1	1		3		1	3			4		32	14	2.29
E	1			3	2		1	1	1	4		1		6	2	3	3					1		29	13	2.23
F	1		2	2	1		1		4		3		2		1		2		1		2	3		25	13	1.92
G	1		3		1			2	1	2		1	1	1	1	1		1		3	3			22	14	1.58
H		5		9	1			2	2			2	2			4	1	5	1	7	3		1	45	14	3.22
I			2		1		3		2				2	3	1	1	2	2	1		3		5	28	13	2.15
J	5	3			3	1			1	1	2		2		3		2	2	1	2	3	1	1	33	16	2.07
K		1			2		3	2		3		3	3	2	2	1		1	1	1		1		26	14	1.86
L		5		4	2	3	1									2	2			1	1			21	9	2.34
M		1				4	5		1	4	4	1	1	3	3	3		3		4	2	2	3	44	16	2.75
N	5		2	1	1	1		1		5	2			2				1	1	2	1	3	1	29	15	1.94
O			2	1	1			5	2		8		2		1	1		1	1				1	26	12	2.17
P		5	4			3		1		2	2	2		1	1		1	4	2	1	1	1	1	32	16	2.00
Q			4	1			4		1			6	2		2	1	3		3	2	1	1	1	32	14	2.29
R		3		3	3	1	2		2	3		1		4	4	1		5		1		2	3	38	15	2.53
S	2	1	1		1	2	1	4	1	3		4	7	1	2			1				8	1	40	16	2.50
T	3	2	5		1	1		1				4	5	1	1	2	2		1		4	1		34	15	2.27
U	1						3	8		1	5	2	1	1	2	7	3		3	1	6		1	45	15	3.00
V	2				1	3			4	1	1	3		1				1	5				6	28	11	2.55
W	5		3	5		1							4	3		1	1	1	1	1		2	1	29	13	2.24
X	3	2	4	2	6	2	1	1	1		1				7	3		1	2		1		2	39	16	2.44
Y	1		1		4	2			4					3	2		5		1	2	3	2		30	12	2.50
Z		3	1		2	1	7		3	6		2					1		3	3			2	34	12	2.84
																								814		

Average frequency per cipher letter $=\frac{814}{26}=31.3$ occurrences.

TABLE XIV.—*Data for reconstruction of primary alphabet no. 3*

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
H (1)					**Y (2)**					**M (3)**				
A	76	14	−34	−.45	A	59	12	−23	−.39	A	67	14	−25	−.37
B	65	12	−29	−.45	B	48	7	−27	−.56	B	56	12	−20	−.36
C	82	13	−48	−.52	C	65	13	−26	−.40	C	73	17	−22	−.30
D	76	12	−40	−.53	D	59	12	−23	−.39	D	68	16	−20	−.29
E	75	13	−36	−.48	E	58	10	−28	−.48	E	66	13	−27	−.41
F	71	16	−23	−.33	F	54	8	−30	−.56	F	62	13	−23	−.37
G	68	15	−23	−.34	G	51	8	−27	−.53	G	59	10	−29	−.49
I	72	15	−16	−.22	I	55	10	−25	−.46	I	63	14	−21	−.33
J	76	20	−16	−.21	J	59	13	−20	−.34	J	67	14	−25	−.37
K	74	14	−32	−.43	K	57	12	−21	−.37	K	65	15	−20	−.31
L	68	7	−47	−.69	L	51	9	−24	−.47	L	59	6	−41	−.70
M	88	15	−43	−.49	M	71	23	−2	−.03	N	64	9	−37	−.58
N	73	15	−28	−.38	N	56	13	−17	−.30	O	64	13	−25	−.39
O	73	10	−43	−.59	O	56	5	−41	−.73	P	70	13	−31	−.44
P	79	16	−31	−.39	P	62	15	−17	−.28	Q	70	21	−7	−.10
Q	70	18	−25	−.32	Q	62	10	−32	−.52	R	75	12	−39	−.52
R	84	17	−33	−.39	R	67	22	−1	−.02	S	76	16	−28	−.37
S	85	15	−40	−.47	S	68	15	−23	−.34	T	70	21	−7	−.10
T	79	19	−22	−.28	T	62	7	−41	−.66	U	82	31	+11	+.13
U	91	18	−37	−.41	U	74	10	−44	−.60	V	65	10	−35	−.54
V	74	12	−38	−.51	V	57	7	−36	−.63	W	62	8	−38	−.61
W	72	14	−30	−.42	W	55	8	−31	−.56	X	65	13	−26	−.40
X	81	18	−27	−.33	X	57	12	−21	−.37	Z	73	12	−37	−.51
Y	76	23	−7	−.09	Z	65	11	−32	−.49					
Z	82	13	−43	−.53										
					H at two intervals									
					M	88	29	−1	−.01					
					R	81	20	−21	−.26					

TABLE XIV.—*Data for reconstruction of primary alphabet no. 3*—Continued

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
U (4)				
A	73	17	−22	−.30
B	63	17	−12	−.19
C	79	13	−40	−.51
D	73	14	−31	−.43
E	72	11	−39	−.54
F	68	14	−26	−.38
G	65	13	−26	−.40
I	69	11	−36	−.52
J	73	14	−31	−.43
K	71	14	−29	−.41
L	65	7	−44	−.68
N	70	10	−40	−.57
O	70	10	−40	−.57
P	76	14	−34	−.45
Q	76	17	−25	−.33
R	81	13	−42	−.52
S	82	20	−22	−.27
T	76	16	−28	−.37
V	71	11	−38	−.54
W	69	[illegible]	−42	−.61
X	78	8	−54	−.69
Z	79	10	−49	−.62
M at two intervals				
A	65	17	−14	−.22
B	54	16	− 6	−.11
S	74	20	−14	−.19

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
B (5)				
A	47	8	−23	−.49
C	53	0	−26	−.49
D	47	5	−32	−.68
E	46	7	−25	−.54
F	42	7	−21	−.50
G	39	6	−21	−.54
I	43	11	−10	−.23
J	47	12	−11	−.28
K	45	10	−15	−.33
L	39	2	−33	−.85
N	44	7	−23	−.52
O	46	7	−25	−.54
P	50	12	−14	−.28
Q	50	8	−26	−.52
R	55	13	−16	−.29
S	56	11	−23	−.41
T	50	7	−29	−.58
V	45	7	−24	−.53
W	43	7	−22	−.51
X	52	7	−31	−.60
Z	53	10	−23	−.43
U at two intervals				
I	69	21	− 6	−.09
J	73	16	−25	−.34
K	71	14	−29	−.41
P	76	16	−28	−.37
R	81	18	−27	−.38

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
I (6)				
A	49	11	−16	−.33
C	55	8	−31	−.56
D	49	15	− 4	−.08
E	48	10	−18	−.38
F	44	7	−23	−.52
G	41	8	−17	−.42
J	49	8	−25	−.51
K	47	11	−14	−.30
L	41	6	−23	−.56
N	46	13	− 7	−.15
O	46	7	−25	−.54
P	52	12	−16	−.31
Q	52	7	−31	−.60
R	57	13	−18	−.32
S	58	11	−25	−.43
T	52	7	−31	−.60
V	47	6	−29	−.62
W	45	12	− 9	−.20
X	54	8	−30	−.56
Z	55	7	−34	−.62
B at two intervals				
D	47	8	−23	−.49
K	45	7	−24	−.53
N	44	10	−14	−.32
W	43	8	−19	−.44
U at three intervals				
D	73	13	−34	−.47
K	71	13	−32	−.45
N	70	16	−22	−.32
W	69	13	−30	−.44

TABLE XIV.—*Data for reconstruction of primary alphabet no. 8*—Continued

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence	Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
N (7)					C (8)					T (9)				
A	59	13	−20	−.34	A	63	17	−12	−.19	A	65	11	−32	−.49
C	65	22	+1	+.02	D	63	11	−30	−.48	D	65	11	−32	−.49
D	59	8	−35	−.59	E	62	10	−32	−.52	E	64	15	−19	−.30
E	58	9	−31	−.53	F	58	10	−28	−.48	F	60	14	−18	−.30
F	54	13	−15	−.28	G	55	17	−4	−.07	G	57	10	−27	−.47
G	51	7	−30	−.59	J	68	14	−26	−.38	J	65	16	−17	−.26
J	59	16	−11	−.19	K	61	17	−10	−.16	K	63	13	−24	−.38
K	57	11	−24	−.42	L	55	10	−25	−.46	L	57	14	−15	−.26
L	51	12	−15	−.30	O	60	9	−33	−.55	O	62	10	−32	−.52
O	56	12	−20	−.36	P	66	17	−15	−.23	P	68	16	−20	−.30
P	62	16	−14	−.23	Q	66	19	−9	−.14	Q	68	15	−23	−.34
Q	62	12	−26	−.42	R	71	12	−35	−.49	R	73	21	−10	−.14
R	67	17	−16	−.24	S	72	15	−27	−.38	S	74	17	−23	−.31
S	68	11	−35	−.52	T	66	21	−3	−.05	V	63	5	−48	−.76
T	62	12	−26	−.42	V	61	7	−40	−.66	W	61	24	+11	+.18
V	57	10	−27	−.47	W	59	9	−32	−.54	X	70	15	−25	−.36
W	55	7	−34	−.62	X	68	12	−32	−.47	Z	71	10	−41	−.58
X	64	14	−22	−.34	Z	69	15	−24	−.35					
Z	65	14	−23	−.35										
W (10)					B (11)					X (12)				
A	58	9	−31	−.54	A	64	14	−24	−.38	A	64	17	−13	−.20
D	58	6	−40	−.69	D	64	15	−19	−.30	D	64	16	−16	−.25
E	57	13	−18	−.32	E	63	12	−27	−.43	E	63	16	−15	−.24
F	53	8	−29	−.55	F	59	12	−23	−.39	F	59	9	−26	−.44
G	50	6	−32	−.64	G	56	12	−20	−.36	G	56	9	−29	−.52
J	58	15	−13	−.22	J	64	15	−19	−.30	J	64	13	−25	−.39
K	56	12	−20	−.36	K	65	12	−29	−.45	K	65	12	−29	−.45
L	50	14	−8	−.16	L	56	10	−26	−.47	L	56	18	−2	−.04
O	55	6	−37	−.67	O	61	13	−22	−.36	O	61	7	−40	−.66
P	61	13	−22	−.36	P	67	17	−16	−.24	P	67	13	−28	−.42
Q	61	9	−34	−.56	Q	67	13	−28	−.42	Q	67	13	−28	−.42
R	66	22	0	0	S	73	13	−34	−.47	S	73	10	−43	−.59
S	67	8	−43	−.64	V	62	13	−23	−.37	V	62	8	−38	−.62
V	56	6	−38	−.68	X	69	22	−3	−.04	Z	70	16	−22	−.31
X	63	17	−12	−.19	Z	70	12	−34	−.49					
Z	64	11	−31	−.48										

TABLE XIV.—*Data for reconstruction of primary alphabet no. 3*—Continued

L (13)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
A	50	13	−11	−.21
D	50	6	−32	−.64
E	49	7	−28	−.57
F	45	8	−21	−.47
G	42	6	−24	−.64
J	50	10	−20	−.40
K	48	8	−24	−.50
O	47	5	−32	−.68
P	53	12	−17	−.32
Q	53	11	−20	−.38
S	59	8	−35	−.59
V	48	4	−36	−.75
Z	56	8	−32	−.57

A (14)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
D	57	15	−12	−.21
E	56	9	−29	−.52
F	52	14	−10	−.19
G	49	11	−16	−.33
J	57	14	−15	−.25
K	55	12	−19	−.35
O	54	10	−24	−.45
P	60	15	−15	−.25
Q	60	13	−21	−.35
S	66	14	−24	−.37
V	55	16	− 7	−.13
Z	63	15	−18	−.29

L at two intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
D	50	13	−11	−.22
F	45	6	−27	−.60
J	47	7	−26	−.55
K	46	7	−25	−.54
P	48	10	−18	−.38
V	48	9	−21	−.44

D (15)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	56	10	−26	−.47
F	52	12	−16	−.31
G	49	9	−22	−.45
J	57	16	− 9	−.16
K	55	13	−16	−.29
O	54	8	−30	−.56
P	60	11	−27	−.45
Q	60	15	−15	−.25
S	66	10	−36	−.55
V	55	14	−12	−.22
Z	63	24	+ 9	+.14

Z (16)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	56	6	−28	−.50
F	52	11	−19	−.37
G	49	15	− 4	−.08
J	57	12	−21	−.37
K	55	10	−25	−.46
O	54	17	− 3	−.06
P	60	13	−21	−.35
Q	60	10	−30	−.50
S	66	14	−24	−.36
V	55	5	−40	−.73

D at two intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
G	53	17	− 2	−.04
O	58	11	−25	−.43

G (17)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	50	10	−20	−.40
F	46	15	− 1	−.02
J	51	14	− 9	−.18
K	49	8	−25	−.51
O	48	9	−21	−.44
P	54	11	−21	−.39
Q	54	9	−27	−.50
S	60	10	−30	−.50
V	49	7	−28	−.57

F (18)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	52	14	−10	−.19
J	53	9	−26	−.49
K	51	15	− 6	−.12
O	50	6	−32	−.64
P	56	12	−20	−.36
Q	56	8	−32	−.57
S	62	14	−20	−.32
V	51	11	−18	−.35

Z at three intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	50	12	−14	−.28
K	57	17	− 6	−.10
P	62	10	−32	−.52
S	68	23	+ 1	+.05

TABLE XIV.—*Data for reconstruction of primary alphabet no. 3*—Continued

S (19)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	67	14	−25	−.37
J	68	13	−29	−.43
K	66	14	−24	−.37
O	65	14	−23	−.35
P	71	12	−35	−.49
Q	71	11	−38	−.54
V	66	15	−21	−.32

V (20)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	51	7	−30	−.59
J	52	12	−16	−.31
K	50	15	− 5	−.10
O	49	5	−34	−.69
P	55	10	−25	−.46
Q	55	10	−25	−.46

K (21)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	55	10	−25	−.46
J	56	14	−14	−.25
O	53	16	− 5	−.09
P	59	14	−17	−.29
Q	59	13	−20	−.34

O (22)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	54	11	−21	−.39
J	55	7	−34	−.62
P	58	9	−31	−.53
Q	58	12	−22	−.38

F at five intervals

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
E	46	13	− 7	−.15
Q	49	4	−37	−.75

E (23)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
J	57	18	− 3	−.05
P	60	13	−21	−.35
Q	60	9	−33	−.55

J (24)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
P	55	24	+17	+.31
Q	55	13	−16	−.29

P (25)

Letter	Total occurrences	Coincidences	Differences	Indices of coincidence
Q (26)	63	22	+ 3	+.05

The completely reconstructed alphabet 3 is as follows:

Alphabet 3.... R X L A D Z Q F S V K O E J P Q H Y M U B I N C T W

With alphabets 1 and 3 at hand, the partial decipherment of the initial groups of a few messages soon leads to the complete reconstruction of the other three alphabets. For example note what these two alphabets give for the beginning of message 11:

Segments....			11					5					19		
Alphabets....	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Cipher........	S	E	Y	B	Z	M	G	S	O	Z	C	M	P	S	Q
Plain text....	O		S			V		T			N				

12345123451

The word OBSERVATION comes to mind almost at once. This means that we have determined the following values:

Alphabet 2	*Alphabet 4*	*Alphabet 5*
Segment 11, $B_p=E_c$	Segment 11, $E_p=B_c$	Segment 11, $R_p=Z_c$
Segment 5, $A_p=G_c$	Segment 5, $I_p=O_c$	Segment 5, $O_p=Z_c$

New values thus determined are inserted in their correct positions with the result that after a short time alphabets 2, 4, and 5 are completely reconstructed. The five alphabets are found to be as follows:

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

Proof of their correctness may be at once established by deciphering the first few groups of message 1. They are as follows:

MLVXK QNXVD GIRIE IMNEE FEXVP HPVZR UKSER MVQCI etc.
EVENI NGREP ORTSS HOULD INCLU DEDAI LYLOS SESSE etc.

"Evening reports should include daily losses . . . "

Once the primary alphabets have been reconstructed, the solution of subsequent messages, written by means of them but employing a wholly different numerical key, is very easy; for the cryptanalyst has only to set the first 5 letters of any message in one segment and look for the plain-text beginning of the message in some other segment. Once the starting point has been found, the number corresponding to the serial number of the message can be inserted in its proper position, and all other numbers of the key determined in a similar manner.

This cipher is of interest also in view of its striking similarity to the Bazeries Disk Cipher, or the "Star Cipher", described in a previous paper [1] by the author.

In the latter cipher, instead of only 5 primary alphabets, there are 20 to 25, and instead of having a numerical key to determine where the cipher equivalents are to be taken, no key of that nature is used, or needed, since the correct line, or generatrix, is easily found because it is the only one that gives intelligible text throughout its length. The segments in the Vogel Cipher correspond to the generatrices in the Star Cipher. In the former, the cipher letters are taken from a definite cycle of generatrices, and not all generatrices are used; in the latter, no such cycle obtains, for it is unnecessary, and all 25 generatrices may be used at random.

The fatal defect in this cipher, from the point of view of its practical indecipherability, is that the segments, or generatrices, are used in a definite sequence or cycle, giving rise to an internal period within messages and an external period between successive messages. It was through them that the recovery of the numerical key and the primary alphabets was possible.

Addendum

The method presented above for determining the index of coincidence will in most cases lead to the correct fitting of two frequency distributions. Since writing the foregoing description of the method for determining the index of coincidence, the author has applied another method which, although it involves additional calculations, will lead to even more accurate results. Since cases which may necessitate a greater refinement in method than that elucidated above will arise, it may be worth while to present the more detailed method.

This method more closely approximates the actual methods used by statisticians and biometricians in their analysis of data, in that it involves the squares of deviations or differences between corresponding observed values.

[1] *Several Machine Ciphers and Methods for their Solution.* Riverbank Publication No. 20. (1918.)

Let us consider the Y sequence of frequencies of table VIII. We desire to find that distribution which, when shifted one interval to the left of the Y distribution, will most closely approximate the Y distribution; i.e., comes from the same "parent" distribution. Returning again to figures 13 and 14, we have the following:

FIGURE 25

A			2	1	2	2	1	1	2	1				1			2	2		1		2	2	
	Y		5	1	1		4					2	2	2			2	6			2		7	2

Pearson [1] gives, in his discussion concerning "Testing Goodness of Fit", the method for determining the probability that random sampling would lead to as large or larger deviations between observations than those observed.

We shall here summarize the paper referred to. Let the samples be given by the frequencies in the same class as follows:

		Total
First sample:	$f_1, f_2, f_3, \quad . \quad . \quad . \quad f_{n'}$	N
Second sample:	$f'_1, f'_2, f'_3, \quad . \quad . \quad . \quad f'_{n'}$	N'

where the totals N and N' differ as little or as widely as we please. The value χ^2 given by $\chi^2 = \frac{1}{NN'}\sum_{i=1}^{n'}\frac{(N'f_i - Nf'_i)^2}{f_i + f'_i}$ is calculated. Tables[2] have been calculated which give, for various values of n', the probability of obtaining a value of χ^2 as big or bigger than the one calculated. In matching alphabets we must be careful to choose the proper value of n' in the tables. It is known that not all 26 letters will appear in a monoalphabet until the total number of letters is about 500. The number of different letters corresponding to various total letters per monoalphabet can be determined, and a curve plotted that will enable one to obtain this value readily.[2] The procedure in applying the method for matching is as follows:

1. The value of χ^2 is calculated as above.
2. The number of different letters, n', corresponding to a total of $N+N'$ letters is obtained from the curve.
3. In the tables for χ^2 the column n' gives the probability corresponding to the calculated value of χ^2.

This probability expresses the chance of getting another set to yield a value of χ^2 at least as large as the one obtained; in other words, the closeness with which the two alphabets match.

Let us proceed to find the value of P for the two distributions shown in figures 13 and 14. We first find the differences between the weighted frequencies in the respective superimposed sequences and then square the differences. Thus:

[1] Pearson, Karl. *On the probability that two independent distributions of frequency are really samples from the same population.* Biometrika, vol. 8 (1911), pp. 250ff.

[2] See, S. Kullback, loc. cit., Table IV and chart 1. The table and chart are reproduced on pages 61-64 of this paper.

FIGURE 26

																							Total
	A		2	1	2	2	1	1	2	1				1		2	2	1		2	2		22
		Y	5	1	1		4					2	2	2		2	6		2		7	2	36
Weighted frequencies			72	36	72	72	36	36	72	36				36		72	72	36		72	72		
			110	22	22		88					44	44	44		44	132		44		154	44	
Differences			38	14	50	72	52	36	72	36		44	44	8		28	60	36	44	72	82	44	
Squares of differences			1,444	196	2,500	5,184	2,704	1,296	5,184	1,296		1,936	1,936	64		784	3,600	1,296	1,936	5,184	6,724	1,936	

Next, we must divide each difference squared by the sum of the corresponding *original* frequencies. Thus:

TABLE XV

Square of difference	Sum of original frequencies	Quotient	Square of difference	Sum of original frequencies	Quotient
1, 444	7	206. 3	64	3	21. 3
196	2	98. 0	0	0	0. 0
2, 500	3	833. 3	0	0	0. 0
5, 184	2	2, 592. 0	784	4	196. 0
2, 704	5	540. 8	3, 600	8	450. 0
1, 296	1	1, 296. 0	0	0	0. 0
5, 184	2	2, 592. 0	1, 296	1	1, 296. 0
1, 296	1	1, 296. 0	1, 936	2	968. 0
0	0	0. 0	5, 184	2	2, 592. 0
1, 936	2	968. 0	6, 724	9	747. 1
1, 936	2	968. 0	1, 936	2	968. 0

To find χ^2 we must now add these quotients and divide the sum by $N \times N'$ or 22×36. The sum of the quotients, 18,628.8, when divided by 792 gives $\chi^2 = 23.5$. In this particular case the further refinement was not applied. Since there are 22 segments we look up the table for $\chi^2 = 23.5$ and $n' = 22$ and find $P = .32$ which means that in 32 of 100 cases we would get by chance a match as bad as or worse than that observed.

We would ordinarily then go through exactly the same steps for all the other sequences in our table VIII against the Y sequence; the sequence which gives the greatest value for P would be the correct one and would indicate which letter follows Y in alphabet 1. But it will be unnecessary here to go through all the calculations since it is desired only to show that this method will give more trustworthy results than the method of coincidences and noncoincidences presented above, and we shall, therefore, show only a few cases.

Let us take only those five sequences in table VIII which when tested against the Y sequence by the previous method gave the greatest indices of coincidence, viz, the P, Q, N, L, and J sequences which gave indices of coincidence of —.12, —.23, —.24, —.28, and —.29, respectively. The data for these sequences upon the method of squares are as follows:

TABLE XVI

																								Total	
P		1	2	1		2			1		1	3			2	1	6			2				22	$\chi^2 = 15.8$ $P = .78$
	Y	5	1	1		4					2	2	2			2	6			2		7	2	36	
Q		4				4	3	2			1		6			1	1		1	2			4	29	$\chi^2 = 23.3$ $P = .32$
	Y	5	1	1		4					2	2	2			2	6			2		7	2	36	
N		3	1			1	3			3		1	2		3	1	1		2	1		4	1	27	$\chi^2 = 21.2$ $P = .45$
	Y	5	1	1		4					2	2	2			2	6			2		7	2	36	
L			2	1		1		4	3			1	3	3		1	2		5			6	1	33	$\chi^2 = 28.9$ $P = .11$
	Y	5	1	1		4					2	2	2			2	6			2		7	2	36	
J		4	1	1				3	1	1	3	2	1			4		3			1	1	1	27	$\chi^2 = 26.4$ $P = .19$
	Y	5	1	1		4					2	2	2			2	6			2		7	2	36	

The values of P (index of probability or coincidence) show that the probability that the P sequence belongs one interval to the left of the Y sequence is a little more than 1.7 times as great as that of the nearest rival to the P sequence, viz, the N sequence. The result is quite gratifying.

We shall show another case. For this we shall choose a case in which a tertiary trial had to be made on the previous method—that is, the method in which a trial with sequences only one interval and two intervals removed failed to give definite and clear-cut results necessitating a third test with a sequence three intervals removed. Note the difficulty in finding the letter which follows K in table VIII. Here even the second trial failed to show whether D, H, Q, or Z follows K. Let us test these sequences against the K sequence using the method of squares. The data are as follows:

TABLE XVII

																										Total	
D		2	2	1	1	5	1	2	1	4	3		3				1		2	2				1	4	35	$\chi^2=19.4$ $P=.55$
	K				3	5	1	1		2	1		6			4	2		4	5	1	1			3	39	
H		1	2	1		3		1		1	4	1	1			5	1		5			4			2	32	$\chi^2=21.7$ $P=.42$
	K				3	5	1	1		2	1		6			4	2		4	5	1	1			3	39	
Q		4				4	3	2			1		6				1	1		1	2				4	29	$\chi^2=21.7$ $P=.42$
	K				3	5	1	1		2	1		6			4	2		4	5	1	1			3	39	
Z			1		4	2	1			2	2						5		2	5	2				6	32	$\chi^2=17.7$ $P=.65$
	K				3	5	1	1		2	1		6			4	2		4	5	1	1			3	39	

The difference between D and Z is not conclusive enough to decide definitely yet, but H and Q can be eliminated at once. Testing for interval 2 will show that D is the proper letter.

It is to be noted that for our purposes it is not even necessary to find the actual value of P; for given n', constant in value, P increases as χ^2 decreases, not in a regular manner it is true but approximately so. Therefore, if n' is constant throughout a series of calculations, the index of coincidence in our case will vary in a general manner inversely with the value of χ^2, the lowest value of χ^2 indicating the greatest degree of probability or the greatest index of coincidence.

It is believed that this method of squares will be found exceedingly valuable in many other cases where considerable refinement of method is necessary to produce clear-cut results in fitting frequency distributions to one another.

TABLE IV.[1] *Test for Goodness of Fit. Values of P.*

χ^2	$n'=3$	$n'=4$	$n'=5$	$n'=6$	$n'=7$	$n'=8$	$n'=9$	$n'=10$	$n'=11$
1	·606531	·801253	·909796	·962566	·985612	·994829	·998249	·999438	·999828
2	·367879	·572407	·735759	·849146	·919699	·959840	·981012	·991468	·996340
3	·223130	·391625	·557825	·699986	·808847	·885002	·934357	·964295	·981424
4	·135335	·261464	·406006	·549416	·676676	·779778	·857123	·911413	·947347
5	·082085	·171797	·287298	·415880	·543813	·659963	·757576	·834308	·891178
6	·049787	·111610	·199148	·306219	·423190	·539750	·647232	·739919	·815263
7	·030197	·071897	·135888	·220640	·320847	·428880	·536632	·637119	·725444
8	·018316	·046012	·091578	·156236	·238103	·332594	·433470	·534146	·628837
9	·011109	·029291	·061099	·109064	·173578	·252656	·342296	·437274	·532104
10	·006738	·018566	·040428	·075235	·124652	·188573	·265026	·350485	·440493
11	·004087	·011726	·026564	·051380	·088376	·138619	·201699	·275709	·357518
12	·002479	·007383	·017351	·034787	·061969	·100558	·151204	·213308	·285057
13	·001503	·004637	·011276	·023379	·043036	·072109	·111850	·162607	·223672
14	·000912	·002905	·007295	·015609	·029636	·051181	·081765	·122325	·172992
15	·000553	·001817	·004701	·010363	·020256	·036000	·059145	·090937	·132061
16	·000335	·001134	·003019	·006844	·013754	·025116	·042380	·066881	·099632
17	·000203	·000707	·001933	·004500	·009283	·017396	·030109	·048716	·074364
18	·000123	·000440	·001234	·002947	·006232	·011970	·021226	·035174	·054964
19	·000075	·000273	·000786	·001922	·004164	·008187	·014860	·025193	·040263
20	·000045	·000170	·000499	·001250	·002769	·005570	·010336	·017913	·029253
21	·000028	·000105	·000317	·000810	·001835	·003770	·007147	·012650	·021093
22	·000017	·000065	·000200	·000524	·001211	·002541	·004916	·008880	·015105
23	·000010	·000040	·000127	·000338	·000796	·001705	·003364	·006197	·010747
24	·000006	·000025	·000080	·000217	·000522	·001139	·002292	·004301	·007600
25	·000004	·000016	·000050	·000139	·000341	000759	·001554	·002971	·005345
26	·000002	·000010	·000032	·000090	·000223	·000504	·001050	·002043	·003740
27	·000001	·000006	·000020	·000057	·000145	·000333	·000707	·001399	·002604
28	·000001	·000004	·000012	·000037	·000094	·000220	·000474	·000954	·001805
29	·000001	·000002	·000008	·000023	·000061	·000145	·000317	·000648	·001246
30	·000000	·000001	·000005	·000015	·000039	·000095	·000211	·000439	·000857
40	·000000	·000000	·000000	·000000	·000001	·000001	·000003	·000008	·000017
50	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000
60	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000
70	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000

[1] Copied from *Tables for Statisticians and Biometricians*, Edited by Karl Pearson, Part I, 2nd Ed., Cambridge University.

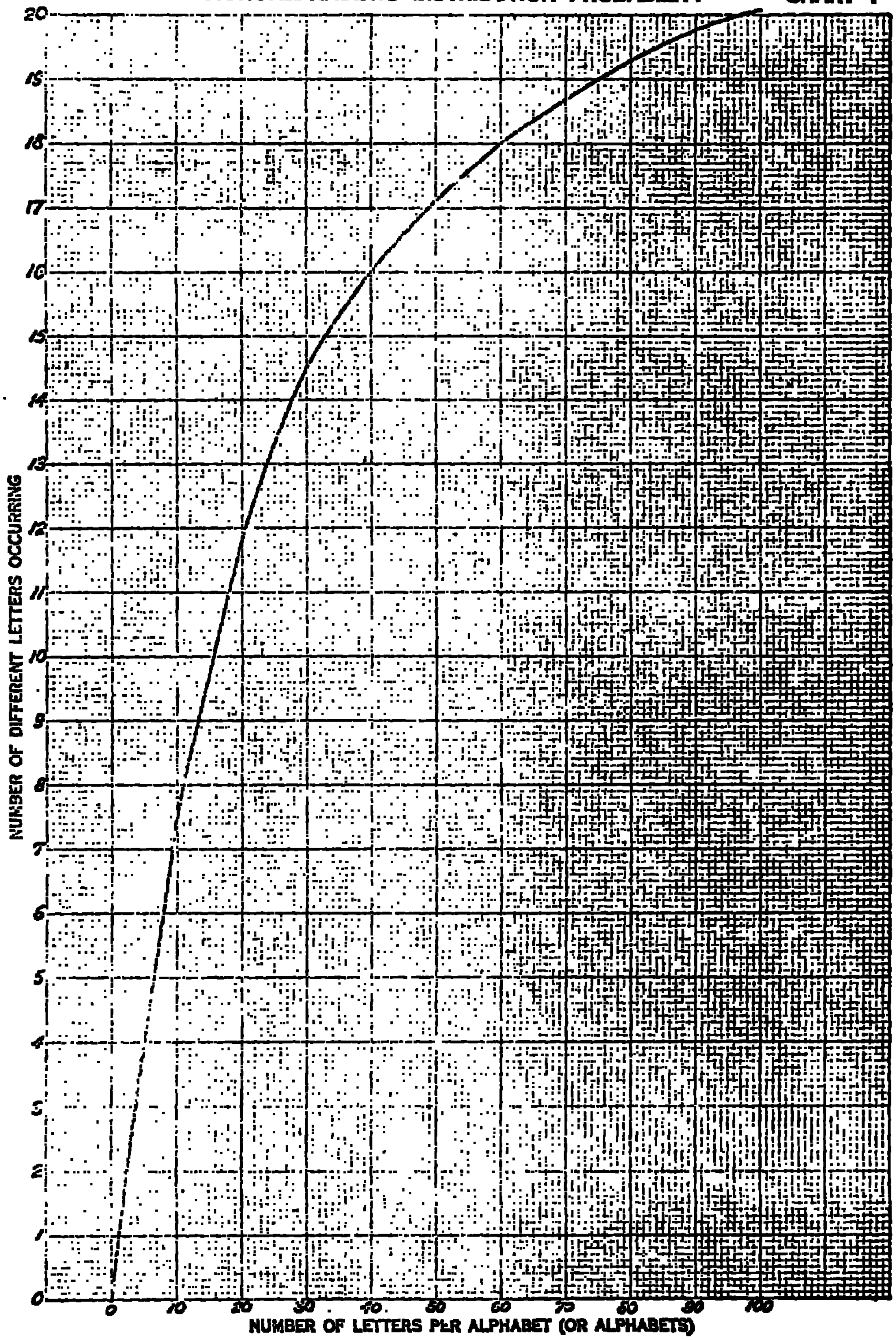
TABLE IV—(*continued*).

χ^2	$n'=12$	$n'=13$	$n'=14$	$n'=15$	$n'=16$	$n'=17$	$n'=18$	$n'=19$	$n'=20$
1	·999950	·999986	·999997	·999999	1·	1·	1·	1·	1·
2	·998496	·999406	·999774	·999917	·999970	·999990	·999997	·999999	1·
3	·990726	·995544	·997934	·999074	·999593	·999830	·999931	·999972	·999989
4	·969917	·983436	·991191	·995466	·997737	·998903	·999483	·999763	·999894
5	·931167	·957979	·975193	·985818	·992127	·995754	·997771	·998860	·999431
6	·873365	·916082	·946153	·966491	·979749	·988095	·993187	·996197	·997929
7	·799073	·857613	·902151	·934711	·957650	·973260	·983549	·990125	·994213
8	·713304	·785131	·843601	·889327	·923783	·948867	·966547	·978637	·986671
9	·621892	·702931	·772943	·831051	·877517	·913414	·940261	·959743	·973479
10	·530387	·615960	·693934	·762183	·819739	·866628	·903610	·931906	·952946
11	·443263	·528919	·610817	·686036	·752594	·809485	·856564	·894357	·923839
12	·362642	·445680	·527643	·606303	·679033	·743980	·800136	·847237	·885624
13	·293326	·369041	·447812	·526524	·602298	·672758	·736186	·791573	·838571
14	·232993	·300708	·373844	·449711	·525529	·598714	·667102	·729091	·783691
15	·182498	·241436	·307354	·378154	·451418	·524638	·595482	·661967	·722598
16	·141130	·191236	·249129	·313374	·382051	·452961	·523834	·592547	·657277
17	·107876	·149597	·199304	·256178	·318864	·385597	·454366	·523105	·589868
18	·081581	·115691	·157520	·206781	·262666	·323897	·388841	·455653	·522438
19	·061094	·088529	·123104	164949	·213734	·268663	·328532	·391823	·456836
20	·045341	·067086	·095210	·130141	·171932	·220220	·274229	·332819	·394578
21	·033371	·050380	·072929	·101632	·136830	·178510	·226291	·279413	·336801
22	·024374	·037520	·055302	·078614	·107804	·143191	·184719	·231985	·284256
23	·017676	·027726	·041677	·060270	·084140	·113735	·149251	·190590	·237342
24	·012733	·020341	·031130	·045822	·065093	·089504	·119435	·155028	·196152
25	·009117	·014822	·023084	·034506	·049943	·069824	·094710	·124915	·160542
26	·006490	·010734	·017001	·025887	·038023	·054028	·074461	·099758	·130189
27	·004595	·007727	·012441	·019254	·028736	·041483	·058068	·078995	·104653
28	·003238	·005532	·009050	·014228	·021569	·031620	·044938	·062055	·083428
29	·002270	·003940	·006546	·010450	·016085	·023936	·034526	·048379	·065985
30	·001585	·002792	·004710	·007632	·011921	·018002	·026345	·037446	·051798
40	·000036	·000072	·000138	·000255	·000453	·000778	·001294	·002087	·003272
50	·000001	·000001	·000003	·000006	·000012	·000023	·000042	·000075	·000131
60	·000000	·000000	·000000	·000000	·000000	·000001	·000001	·000002	·000004
70	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000	·000000

TABLE IV—(*continued*)

χ^2	$n'=21$	$n'=22$	$n'=23$	$n'=24$	$n'=25$	$n'=26$	$n'=27$	$n'=28$	$n'=29$	$n'=30$
1	1·	1·	1·	1·	1·	1·	1·	1·	1·	1·
2	1·	1·	1·	1·	1·	1·	1·	1·	1·	1·
3	·999996	·999998	·999999	1·	1·	1·	1·	1·	1·	1·
4	·999954	·999980	·999992	·999997	·999999	1·	1·	1·	1·	1·
5	·999723	·999868	·999939	·999972	·999987	·999994	·999998	·999999	1·	1·
6	·998598	·999427	·999708	·999855	·999929	·999966	·999984	·999993	·999997	·999999
7	·996085	·998142	·999080	·999452	·999711	·999851	·999924	·999962	·999981	·999991
8	·991868	·995143	·997160	·998371	·999085	·999494	·999726	·999853	·999924	·999960
9	·982907	·989214	·993331	·995957	·997595	·998596	·999194	·999546	·999748	·999863
10	·968171	·978912	·986304	·991277	·994547	·996653	·997981	·998803	·999302	·999599
11	·946223	·962787	·974749	·983189	·989012	·992946	·995549	·997239	·998315	·998988
12	·916076	·939617	·957379	·970470	·979908	·986567	·991173	·994294	·996372	·997728
13	·877384	·908624	·933161	·951990	·966121	·976501	·983974	·989247	·992900	·995384
14	·830496	·869599	·901479	·926871	·946650	·961732	·973000	·981254	·987189	·991377
15	·776408	·822952	862238	·894634	·920759	·941383	·957334	·969432	·978436	·985015
16	·716624	·769650	·815886	·855268	·888076	·914828	·936203	·952947	·965819	·975536
17	·652974	·711106	·763362	·809251	·848662	·881793	·909083	·931122	·948589	·962181
18	·587408	·649004	·705988	·757489	·803008	·842390	·875773	·903519	·926149	·944272
19	·521826	·585140	·645328	·701224	·751990	·797120	·836430	·870001	·898136	·921288
20	·457930	·521261	·583040	·641912	·696776	·746825	·791556	·830756	·864464	·892927
21	·397132	·458944	·520738	·581087	·638725	·692609	·741964	·786288	·825349	·859149
22	·340511	·399510	·459889	·520252	·579267	·635744	·688697	·737377	·781291	·820189
23	·288795	·343979	·401730	·460771	·519798	·577564	·632947	·685013	·733041	·776543
24	·242392	·293058	·347229	·403808	·461597	·519373	·575965	·630316	·681535	·728932
25	·201431	·247164	·297075	·350285	405760	·462373	·518975	·574462	·627835	·678248
26	·165812	·206449	·251682	·300866	·353165	·407598	·463105	·518600	·573045	·625491
27	·135264	·170853	·211226	·256967	·304453	·355884	·409333	·463794	·518247	·571705
28	·109399	·140151	·175681	·215781	·260040	·307853	·358458	·410973	·464447	·517913
29	·087759	·114002	·144861	·180310	·220131	·263916	·311082	·360899	·412528	·465066
30	·069854	·091988	·118464	·149402	·184752	·224289	·267611	·314154	·363218	·414004
40	·004995	·007437	·010812	·015369	·021387	·029164	·039012	·051237	·066128	·083937
50	·000221	·000365	·000586	·000921	·001416	·002131	·003144	·004551	·006467	·009032
60	·000007	·000013	·000022	·000038	·000064	·000104	·000168	·000264	·000407	·000618
70	·000000	·000000	·000001	·000001	·000002	·000004	·000007	·000011	·000019	·000030

MONOALPHABETIC DISTRIBUTION PROBABILITY CHART 1



Part II

THE SCHNEIDER CIPHER[1]

Details of Encipherment

This system aims to make the entire operation of encipherment and decipherment dependent upon the knowledge of a single key word agreed upon in advance by the correspondents. Let us suppose the word to be T R E B I Z O N D.

An arbitrarily mixed alphabet is derived from a generating rectangle based upon this key word, with a slight departure from the usual method. Instead of inscribing the letters in the generating rectangle in the key-word sequence, as is usual, the initial letter of the second line may be any letter except one in the key word itself. The remaining letters then follow in the key-word sequence. Thus suppose we choose K as this initial letter for the second line. Our rectangle is constructed in this manner:

Figure 27

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

By taking the columns in succession and writing them in two lines of 13 letters each we have the following:

Figure 28

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

These two lines constitute alphabet 1, in which T=Q, K=F, X=Z, etc. The values are all reciprocal.

All the other alphabets are derived from alphabet 1 by the simple expedient of carrying the upper half of alphabet 1 to the third line and revolving the sequence *one* letter to the right. Thus:

Figure 29

```
Alphabet 1 { (1)  T K X R L Y E M A B P C I
           { (2)  Q F Z S G O U H N V J D W } Alphabet 2
             (3)  I T K X R L Y E M A B P C }
```

The juxtaposition of lines 2 and 3 results in the formation of alphabet 2, in which Q=I, F=T, Z=K, etc., also completely reciprocal.

---

[1] Schneider, L. *Description d'un système cryptographique à l'usage de l'armée.* Paris, 1912. This cipher was submitted for examination by Maj. Otto Holstein, M.I.-Res. who furnished the writer, who was then at the Riverbank Laboratories, with a translation of Commandant Schneider's paper. Attempts to locate in American libraries and bookstores the work cited to see if the inventor had offered a sample message for solution proved of no avail; fortunately, a copy of the paper was sent in quite accidentally by an obscure book dealer shortly after the first draft of this manuscript was prepared. No sample message is given. Upon my request an assistant prepared the message which will presently be given, and the solution of which was attained by the principles to be elucidated.

Alphabet 3 is constructed by moving line 2 to line 4, revolving the sequence *two* letters to the right. Thus:

FIGURE 30

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

The juxtaposition of lines 3 and 4 results in the formation of alphabet 3, in which I=D, T=W, K=Q, etc.

Continuation of this process can result in the production of a total of 13 different secondary reciprocal alphabets. As a rule only a limited number of these secondary alphabets are employed, usually not to exceed the first 6 or 7.

One of the main features of the method of encipherment is that groups of unequal lengths are enciphered in cyclic fashion by means of several alphabets. That is, the text is broken up into groups containing 1, 2, 3, . . . letters enciphered by different alphabets, the groups being repeated in sets, as explained below.

Let us, as one of the correspondents, determine that the cycle is to consist of six groups. Taking the first 6 letters of line 1, that is the upper half of alphabet 1, their numerical values on the basis of their relative positions in the normal or straight alphabet, are as follows:

T	K	X	R	L	Y
4	1	5	3	2	6

This sequence of numbers, 4–1–5–3–2–6, constitutes a cycle designated hereafter as the "Interruption key." It partakes of the nature of an "interrupter" in that it dictates the number of letters in each of the groups of irregular length treated in encipherment. Thus, the first group contains 4 letters; the second, 1 letter; the third, 5 letters, etc. The seventh group would begin a repetition of the cycle, and it would contain 4 letters; the eighth group, 1 letter, etc. This cycle is repeated many times within a message.

Encipherment within each group is regular in the succession of the alphabets employed. Thus, if three alphabets are determined upon by the correspondents, the alphabets would be employed in the following sequence in the interruption key given above:

FIGURE 31

Length of group	4	1	5	3	2	6
Alphabets	1 2 3 1	1	1 2 3 1 2	1 2 3	1 2	1 2 3 1 2 3

Note that in groups containing more letters than the number of alphabets decided upon, the sequence of alphabets is repeated. Thus, in the 5-letter group, we have the sequence of alphabets 1–2–3–1–2; in the 6-letter group, 1–2–3–1–2–3. Encipherment then proceeds by alphabets according to the distribution of numbers within the groups. For example, group 1 containing the sequence of numbers 1–2–3–1, the first letter is enciphered by means of alphabet 1; the second, by alphabet 2, and the third, by alphabet 3. The fourth letter, however, is again enciphered by alphabet 1.

After encipherment, the order of the cipher letters within the groups is reversed throughout the message. Thus, suppose that the encipherment using the three alphabets and key above were as follows:

FIGURE 32

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

The letters within the cipher groups are then reversed yielding the following:

FIGURE 33

HOMU O FADXW BHS ED EONYMW UTRA R XAZHQ VLG

It is now necessary that the information necessary to decipher the message be conveyed to the recipient, who, of course, is already in possession of the key word. This information is transmitted in the form of an "Indicator group", consisting of 3 letters, whose location within the message has been previously determined in a manner to be explained presently. The first letter of the indicator group gives the initial letter of the second line of the generating rectangle; the second, the number of alphabets employed; and the third, the length of the interruption key, from which its sequence can be derived from the upper half of alphabet 1, as already explained.

In the case above, the initial letter is K. The number of alphabets being three, the third letter in the upper half of alphabet 1, viz, X, forms the second letter of the indicator group. The length of the interruption key being six, the sixth letter of the upper half of alphabet 1, viz, Y, forms the third letter of the indicator group. The knowledge of this letter enables the recipient to construct the interruption key. The indicator group for the message above is, therefore, KXY.

This indicator group is then inserted in the cipher text in a position determined by a previously agreed upon letter of the key word, usually either the first or the last letter. Thus, if the correspondents agree upon the first letter of the key word, this indicator group would be inserted after the twentieth letter of the cipher text, because T, the first letter of the key word, occupies the twentieth position in the normal or straight alphabet. The cipher message above would read, therefore, as follows:

MESSAGE

HOMUO FADXW BHSED EONYM KXYWU TRARX AZHQV etc.

By varying the elements of the indicator group, a great number of combinations can be obtained from one key word. Schneider says:

> With a key of 6 letters, if one sets out to modify the table of alphabets, one obtains only 26 − 6 = 20 different cryptograms from the same text.
>
> If, in addition to the table of alphabets, one modifies the number of alphabets employed and the number of letters to form the numerical key, assuming that one employs 3, 4, 5, or 6 alphabets (a total of 4 combinations) and 5, 6, 7, or 8 letters to form the numerical key (also a total of 4 combinations), one would obtain 20 × 4 × 4 = 320 different cryptograms from one and the same text.

We shall first elucidate the principles by means of which a single message may be solved, and then proceed to the solution of a series of messages in the same key.

1. **Solution of a single message.**—Given the following message:

```
KHNVL  IQKGK  NNHKV  QEEXK  XYXOP  MSIEE  TPDKU  LISYZ  HWBRE  SATHR
KZRGM  NJGKD  QKVVM  FQBKE  NEIHA  EAAME  KHFLW  XRKEO  KMHFM  WAFTW
ESPEB  DDGWP  JPXGD  ZVWUX  LZAYU  ENILH  AIUUA  EABRE  PWKFV  JJKIP
EMENF  SBVYZ  KWDMK  VLODO  OBFMB  BYWOQ  YVKVI  XDGIK  HEONE  EIWKW
PAECL  RNIHN  NMODQ  NIKAO  JKEOZ  TCFWD  JJODW  ZUAES  ZDWKG  KKBDT
XNKGD  IHTKS  NPGDU  RQFMD  ONAZA  ZMWCL  RHEOF  ZWFHV  KZYGE  VPNPK
AQAHM  DCCKK  UGDJW  ILIAB  ECESG  SPFEP  KOPIS  HHODN  DMKFI  SYQMZ
KGIQQ  KTWYG  JWNPK  NWBHE  ULJVQ  ZZWIM  LWWNJ  UCDQQ  KKFDG  WPMFS
KIEBC  CDXLF  MDODO  EANKH  MHFZT  CPMPP  WDAVL  MPXKR  OJKFN  MVPRA
OEWJO  JVVWR  MDEWA  FYWHX  ENMEN  APRNP  VCUCL  HYFMN  LKHVP  NOJKF
WWCLR  AOOWZ  VJVVW  RMDEW  AFYWO  XENGZ  FKVWP  XUNUM  AYMGX  VNKOJ
KFSVB  ZRNLE  IONGM  NVFTD  HGBJO  LIPOW  NPPWK  NREPM  DWXYE  XNZON
UEVOP  WUQQQ  JNMPV  FQJWN  KKOLI  POWKO  VBHFY  MEPMD  AFMHF  ABEFD
MKHWU  EODRL  ISYDK  VZKWK  MIQXK  THYNS  VHEQA  AEVIX  PZTQK  IAMFT
TALOV  JKVHD  UETCL  MHVWR  HFUNN  GFCKK  ESEYE  JUPWW  BC
```

Principles of Solution

This cipher belongs to the class of combined substitution-transposition methods. The substitution, however, does not follow the method of the ordinary periodic multiple-alphabet system, nor is the transposition regular or geometrical in its nature. We shall consider first the results of the method of substitution.

While at first glance it may be thought that in this system, the encipherment of the groups of different lengths eliminates the regularity or cyclic nature of the ordinary periodic multiple-alphabet cipher, such is not strictly the case. It is true that the groups are irregular in length, due to the action of the interrupter, but at the same time, since the interruption key repeats itself many times throughout a message, *there will be regular and cyclic repetitions of constant sets of groups*. In other words, there is a cycle in the system, composed of a constant number of groups of irregular lengths. We shall first proceed to determine the length of this cycle.

The length of this cycle is dependent upon the length of the interruption key, and is the sum of the constituent numbers of the key. It is obvious that the key can be 2, 3, 4 . . . up to 13 numbers in length. It cannot be more than 13 because of the manner in which the key is derived from the upper half of alphabet 1. (See p. 66.)

Now the numbers in these interruption keys do not repeat themselves, and it follows, therefore, as a simple mathematical fact, that the sums of the numbers composing every possible key—in other words, the lengths of the possible cycles—are constant and determinate quantities. Thus, if the key consists of 2 numbers, the cycle will be $1+2=3$ letters in length; if it consists of 3 numbers, the cycle will be $1+2+3=6$ letters in length. The following table gives the length of all possible cycles of the entire system.

TABLE XVIII

Length of key	Length of cycle	Length of key	Length of cycle
2	$1+2=3$	8	$28+8=36$
3	$3+3=6$	9	$36+9=45$
4	$6+4=10$	10	$45+10=55$
5	$10+5=15$	11	$55+11=66$
6	$15+6=21$	12	$66+12=78$
7	$21+7=28$	13	$78+13=91$

In order to determine the correct cycle length we proceed to apply the coincidence method described in the first part of part I of this paper. The message is written out in lines of 3, 6, 10, 15 . . . letters in length to correspond to the various hypotheses regarding length of cycle, as given in table XVIII.

We must first take into consideration, however, the fact that somewhere within the body of the message there is a group of 3 letters, composing the indicators of the message, and since these 3 letters do not form a part of any cycle, they might throw all the cycles succeeding them out of phase with the cycles preceding them. But, from the method given for fixing the location of the indicator group, the latter must occur somewhere within the first 26 letters of the cipher message. By discarding the first 26 letters of our message then, we may be sure that we have eliminated this source of distortion. Starting with the twenty-seventh letter of the message, viz, S, we proceed to set up the message according to various hypothetical cycle lengths. We may omit the lengths 3 and 6, at the start as highly improbable. In this case no table of coincidences is really necessary because the arrangement of the message on the assumption of a cycle of 28 letters gives so many repetitions within the same columns that further corroboration is unnecessary. Note the repetitions of long polygraphs below:

FIGURE 34

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

The next step is to compile individual frequency tables from the columns. There will, of course, be 28 of them. They have been consolidated in table XIX, and the average frequency per column is given.

Table XIX

Column	Cipher letters																										Frequency	Number of different letters	Average frequency
	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			3		2	4				1	2	2	3	2				2	2			2		1			26	12	2.16
2	1		2		3	2		3	2		2	3		2	2	3				1							26	12	2.16
3				1	1	2		1	2	1	4		1			3	2	3		1		1	1	1		1	26	16	1.62
4	4			1	1		2	2		2	2		1	1	3			1		1			3	2			26	14	1.86
5		1	1	4	2	1	1				4	1		2	3	1		1		1	1		2				26	15	1.75
6		2		1	3	2	2		1		1			2	3	2				1	1		3			2	26	14	1.86
7				3	1	3		3			1		1			2	1		1	1		4	3	2			26	13	2.00
8	1		2	2	1					4	3		3	2	1					1			1		3	2	26	13	2.00
9		1			3	1				2	4		1	1	2	1					5	3	2				26	12	2.16
10		1	2			3			1	4	2	3	3						3			2		2			26	11	2.36
11	1							4	4	1	1	4	1	2	1	1				1	1	3			1		26	14	1.86
12				3		2		2	2							5		1	2			4	2		1	2	26	11	2.36
13	2				3		1	1			1		2	1	3		2				2		5		3		26	12	2.16
14		3		2	2	1	2		1				1	1			3	2					3		1	4	26	13	2.00
15	4	2	1		2			2			2		4		1					1	3			1	2	1	26	13	2.00
16	1		4	3	3	1	1			1				3	1			1			1	3	2			1	26	14	1.86
17		1		1	6	1			1		2	3		4	1		1	1		1		2					25	13	1.92
18		1			1	1	1		1		2	2	4		1			1	3			3	3	1			25	14	1.78
19	2	1			2		1	3	5			4	1		2	3										1	25	11	2.27
20				3		4		2		1				1		1	1		4			3	2	3			25	11	2.27
21	3			1			1				2		1		2	5	1				1		3		5		25	11	2.27
22				2	2	2	1		1		3		1	4			1	2		1			3			2	25	13	1.92
23	3	1			2		1	3	2	1	2			1	4		1		1	2	1						25	14	1.78
24	1			2		1				2	3					3	3	1			2	2	2	2		1	25	13	1.92
25	4	1	1	2	2			1		1	6		2		1	1					1				1	1	25	14	1.78
26		1		2	4	2	2	1	1		1		1	4	2								1			3	25	13	1.92
27	4			1			1			2	3		4	2	1	1	1	2					3				25	12	2.08
28		1		1	1	1	2	1	1		2		2	4		2	1			1			3			1	25	16	1.56

Showing frequency distributions for the various columns of figure 34.

Now, according to table XVIII, a cycle length of 28 requires that the interruption key consist of 7 numbers, obviously the digits 1 to 7. We do not know how many alphabets are involved, but they probably do not exceed 7. If only 3 alphabets are involved, then they would be distributed in 7 irregular groups as follows:

TABLE XX.—*Basis of 3 alphabets*

Groups	Alphabets
1	1
2	1—2
3	1—2—3
4	1—2—3—1
5	1—2—3—1—2
6	1—2—3—1—2—3
7	1—2—3—1—2—3—1

Alphabet 1 would be employed twelve times, alphabet 2, nine times, and alphabet 3, seven times. The following data give the number of times the various alphabets would be employed on different hypotheses:

TABLE XXI

Basis of 4 alphabets

Groups	Alphabets
1	1
2	1—2
3	1—2—3
4	1—2—3—4
5	1—2—3—4—1
6	1—2—3—4—1—2
7	1—2—3—4—1—2—3

Frequencies

Alphabet 1, 10 times
Alphabet 2, 8 times
Alphabet 3, 6 times
Alphabet 4, 4 times

Basis of 5 alphabets

Groups	Alphabets
1	1
2	1—2
3	1—2—3
4	1—2—3—4
5	1—2—3—4—5
6	1—2—3—4—5—1
7	1—2—3—4—5—1—2

Frequencies

Alphabet 1, 9 times
Alphabet 2, 7 times
Alphabet 3, 5 times
Alphabet 4, 4 times
Alphabet 5, 3 times

Basis of 6 alphabets

Groups	Alphabets
1	1
2	1—2
3	1—2—3
4	1—2—3—4
5	1—2—3—4—5
6	1—2—3—4—5—6
7	1—2—3—4—5—6—1

Frequencies

Alphabet 1, 8 times
Alphabet 2, 6 times
Alphabet 3, 5 times
Alphabet 4, 4 times
Alphabet 5, 3 times
Alphabet 6, 2 times

Basis of 7 alphabets

Groups	Alphabets
1	1
2	1—2
3	1—2—3
4	1—2—3—4
5	1—2—3—4—5
6	1—2—3—4—5—6
7	1—2—3—4—5—6—7

Frequencies

Alphabet 1, 7 times
Alphabet 2, 6 times
Alphabet 3, 5 times
Alphabet 4, 4 times
Alphabet 5, 3 times
Alphabet 6, 2 times
Alphabet 7, 1 time

We shall now try to determine how many alphabets were employed, and how they are distributed within the cycle. The basis for this determination rests upon the possibility of comparing the various frequency distributions produced by each alphabet and determining which are similar. In other words, we proceed to classify the various frequency tables into several groups, the number of different groups corresponding to the number of alphabets employed in the message. Accordingly, we select one of the frequency tables as the most closely approximating that produced by a single mixed alphabet, and try to find others which seem to be identical with it, as evidenced by the index of coincidence. Since there will be more than one alphabet in each group, we shall look for more than one high index of coincidence. For example, if we find three indices which are much higher than the remaining indices, we may conclude that the alphabets corresponding to these indices are identical with the alphabet which is being used for comparison.

Returning to table XIX, and following the reasoning given on page 37 of the first part of this paper, columns 10 and 12 give the closest approximations to a theoretical single frequency table, and we shall start fitting frequencies with them as bases.

Taking the frequency table for column 10 and applying it to that for every other column, we get the coincidence data shown herewith:

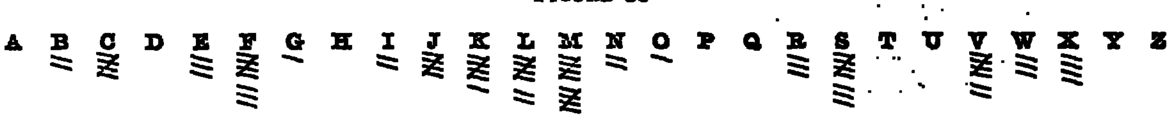
TABLE XXII

Column	1	2	3	4	5	6	7	8	9	11	12	13	14	15
Total occurrences	52	52	52	52	52	52	52	52	52	52	52	52	52	52
Coincidences	17	10	9	7	6	5	10	11	9	8	7	3	5	8
Differences	−1	−22	−25	−31	−34	−37	−22	−19	−25	−28	−31	−43	−37	−28
Indices of coincidence	−.02	−.42	−.48	−.60	−.65	−.71	−.42	−.37	−.48	−.54	−.60	−.83	−.71	−.54

Column	16	17	18	19	20	21	22	23	24	25	26	27	28
Total occurrences	52	51	51	51	51	51	51	51	51	51	51	51	51
Coincidences	6	10	16	6	11	3	6	6	8	7	6	7	8
Differences	−34	−21	−3	−33	−18	−42	−33	−33	−27	−30	−33	−30	−27
Indices of coincidence	−.65	−.41	−.06	−.65	−.35	−.83	−.65	−.65	−.53	−.59	−.65	−.59	−.53

The indices of coincidence for columns 1 and 18 are so high, relative to those for all other columns, that we may conclude at once that the frequency tables for these two columns belong to the same alphabet as column 10. To corroborate this calculation, let us consolidate the three frequency tables to see whether the result of such consolidation will present the appearance of a single mixed substitution alphabet.

FIGURE 35



Consolidated frequency table of columns 1, 10, and 18

We note at once that this frequency distribution gives every indication of being that of a single mixed alphabet.

Returning to the table of indices of coincidence for all other columns tried with column 10, we note that the indices for columns 1 and 18 are in reality so much higher than those for all other columns that we may assume with a fair degree of certainty that the three columns thus classified constitute all the columns applying to one alphabet. Reference to the diagram (table XX) for a basis of 3 alphabets shows that alphabet 1 is used 12 times, alphabet 2, 9 times, and alphabet 3, 7 times. Now if we are correct in our assumption that columns 1, 10, and 18 constitute all the columns for one alphabet, then this assumption automatically rules out the hypothesis that the problem involves 3 alphabets because, on the latter hypothesis, there cannot be an alphabet which is used but three times; and, for the same reasons, a hypothesis of 4 alphabets is eliminated. It is possible, however, to have an alphabet which is used but three times on hypotheses of 5, 6, or 7 alphabets. In each of these cases, the alphabet to which columns 1, 10, and 18 would belong is alphabet 5.

Let us proceed now to find another alphabet to which some of the other frequency tables belong. The table for column 12, having the same average frequency as that for column 10, is taken as a basis for comparison for the next classification of frequency tables. We may omit from the calculation the frequency distributions for columns 1, 10, and 18, since they have already been classified. The data for this test are as follows:

TABLE XXIII

Column	2	3	4	5	6	7	8	9	11	13	14	15
Total occurrences	52	52	52	52	52	52	52	52	52	52	52	52
Coincidences	9	13	6	8	10	16	6	7	9	4	9	4
Differences	−25	−13	−34	−28	−22	−4	−34	−31	−25	−40	−25	−40
Indices of coincidence	−.48	−.25	−.65	−.54	−.42	−.08	−.65	−.60	−.48	−.77	−.48	−.77
Column	**16**	**17**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Total occurrences	52	51	51	51	51	51	51	51	51	51	51	51
Coincidences	11	6	8	15	9	10	5	11	6	9	5	10
Differences	−19	−33	−27	−6	−24	−21	−36	−18	−33	−24	−36	−21
Indices of coincidence	−.37	−.65	−.53	−.12	−.46	−.41	−.70	−.35	−.65	−.46	−.70	−.41

We may with a fair degree of certainty conclude that the frequency distributions for columns 7, 12, and 20 constitute members of another set. But it is impossible in this system to have two alphabets which are used exactly the same number of times; and since we have already assumed that columns 1, 10, and 18 constitute the alphabet which is used three times, we are forced to conclude that one or more additional columns in this test belong with columns 7, 12, and 20. Now the index for column 3, viz, −.25, is the closest to the indices for columns 7 and 20, and it is perhaps likely that it belongs with these columns. But the index for column 24 is −.35, and that for column 16, −.37, so that we must apply a secondary test. Let us consolidate the frequency tables for columns 7, 12, and 20, and then make a calculation of index of coincidence for columns 3, 24, and 16.

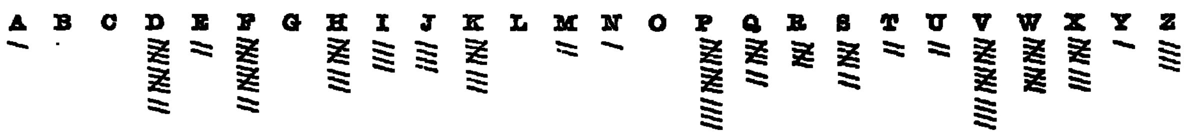
FIGURE 36

Consolidated frequency table of columns 7, 12, and 20

Column	3	16	24
Frequency	103	103	102
Coincidences	21	14	18
Differences	−40	−61	−48
Indices of coincidence	−.39	−.59	−.47

It is very probable that column 16 does not belong in the same alphabet with columns 7, 12, and 20, but columns 3 and 24 do. To corroborate, let us add their corresponding frequency distributions to the consolidated table.

FIGURE 37



Consolidated frequency table of columns 3, 7, 12, 20, and 24

There is every indication that we are dealing with a distribution for a single mixed alphabet, and that no foreign elements have been introduced. If we have succeeded in collecting all the columns which should belong to the alphabet being studied, it follows that the latter must be alphabet 3, for it is used five times on a hypothesis of 5, 6, or 7 alphabets.

Let us now consider the sequence of alphabet numbers which results when we set down the alphabet to which each of the columns now classified belongs.

FIGURE 38

```
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
5 3     3     5  3                5   3            3
```

Now it will be remembered that in encipherment the letters in each group were reversed before grouping into fives for telegraphic transmission. It follows, therefore, that the alphabet numbers applying to columns when the message is set up as shown in figure 32 will be in groups of descending series of numbers and this descent is complete in each case to number 1. Thus, for example, the series in the illustrative message on page 67 are as follows:

FIGURE 39

```
1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21
1-3-2-1-1-2-1-3-2  1- 3- 2- 1- 2- 1- 3- 2- 1- 3- 2- 1
```

Returning now to figure 38, we see that our classification of columns is in accordance with the requirement of descending series of numbers, and we may even fill in many numbers without any further test. Thus, the sequence becomes:

FIGURE 40

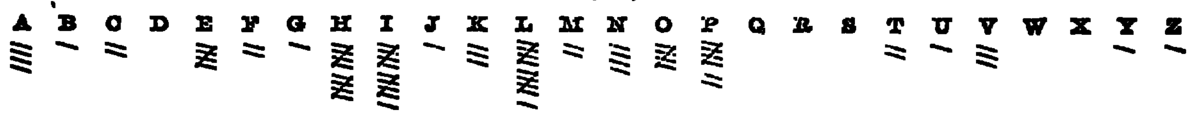
```
1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28
5-4-3-2-1  3-2-1- 5- 4- 3- 2- 1            5- 4- 3- 2- 1    3- 2- 1
```

To corroborate this sequence, let us consolidate the frequency tables for columns 2, 11, and 19 into one table; those for columns 4, 8, 13, 21, and 25 into a second table; and those for columns

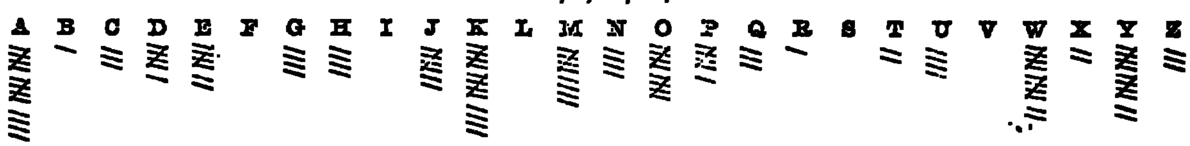
5, 9, 14, 22, and 26 into a third table. We do this to see if the resulting table in each case corresponds to that of a single mixed alphabet.

FIGURE 41

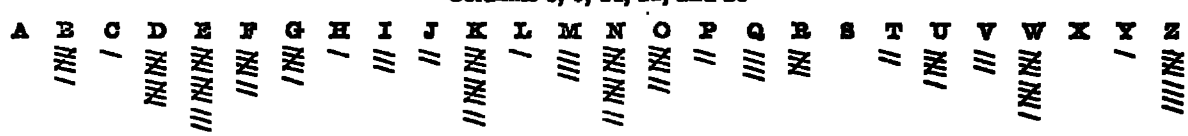
Columns 2, 11, and 19



Columns 4, 8, 13, 21, and 25



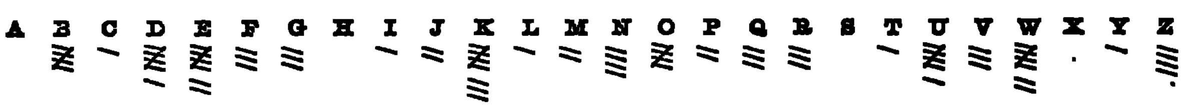
Columns 5, 9, 14, 22, and 26



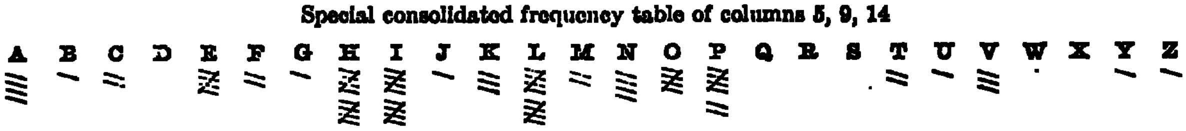
The approximation to single alphabet distributions is, in each case, very close, and we may assume our work thus far to be correct. The only columns remaining to be assigned to alphabets are 6, 15, 16, 17, 23, 27, and 28.

Now column 6 can only belong to either alphabet 4 or alphabet 1; and the same alternatives present themselves with respect to column 23. This is because these columns, being bounded on either side by columns already distributed into alphabets 1 and 3, must either be members of the descending sequence 4 3 2 1, or must be the column forming that part of the interruption key which consists of the number 1, thus constituting the isolated column which must always be enciphered by alphabet 1. In order to determine which of these alternative assignments is the case, we may make a single calculation by finding the index of coincidence when column 6 is grouped with columns 5, 9, 14, 22, and 26, and when grouped with columns 2, 11, and 19. In order to make the results strictly comparable we should include but three columns belonging to alphabet 1, because there are only three columns assigned thus far to alphabet 4. Let us make a special consolidation of columns 5, 9, and 14 for this test.

FIGURE 42



Special consolidated frequency table of columns 5, 9, 14



Consolidated frequency table of columns 2, 11, and 19

Total frequency, columns 6, 5, 9, and 14	104	Total frequency, column 6, 2, 11, and 19	103
Total coincidences, column 6 with 5, 9, and 14	26	Total coincidences, column 6 with 2, 11, and 19	19
Difference 3(26)—104	—26	Difference 3(19)—103	—46
Index of coincidence	—.25	Index of coincidence	—.45

There seems to be no doubt that column 6 belongs to alphabet 1 and not to alphabet 4. This automatically makes it necessary that column 23 be assigned to alphabet 4 since we cannot have two no. 1 columns in the same interruption key.

Let us set down what we have:

FIGURE 43

```
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
5-4-3-2-1-1-3-2-1- 5- 4- 3- 2- 1            5- 4- 3- 2- 1- 4- 3- 2- 1
```

We now proceed to compare these sequences with the sequences of alphabets within the reversed groups on hypotheses of 5, 6, and 7 alphabets.

We may conclude, by a short analysis, that a hypothesis of five alphabets must be correct, because the distribution of columns as now made, if it be correct (and we feel fairly certain of our work so far), allows no room for any alphabet 6 or 7. This analysis is fairly simple. Let us assume a hypothesis of 6 alphabets.

Consider the two places where alphabet 6 would have to fall:

FIGURE 44

```
      13-14-15-16-17-18-19 . . . . . 25-26-27-28-1-2
. . .  2- 1      6- 5- 4-. . . . .  2- 1    6-5-4- . . .
```

According to this diagram, column 27, isolated as it stands, would have to bear the number 1. But we found, already, that column 6 is the column which forms the isolated group containing only one number. Hence, a hypothesis of 6 alphabets cannot be correct.

Let us assume a hypothesis of 7 alphabets. Note these two places once more:

FIGURE 45

```
      13-14-15-16-17-18-19 . . . . . 25-26-27-28-1-2
. . .  2- 1    7- 6- 5- 4- . . . . .  2- 1- 7- 6-5-4- . . .
```

Here again we would have an isolated column (15), which for similar reasons cannot bear the number 1.

We are left, therefore, only the hypothesis of 5 alphabets. This requires that alphabet 1 be used nine times, and alphabet 2, seven times. We have already distributed 6 columns to alphabet 1, and 5 columns to alphabet 2, leaving 3 more columns to be assigned to the former, and 2 more to the latter. We may assign columns 27 and 28 immediately to alphabets 2 and 1, respectively. The sequence thus becomes:

FIGURE 46

```
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
5-4-3-2-1-1-3-2-1- 5- 4- 3- 2- 1            5- 4- 3- 2- 1- 4- 3- 2- 1- 2- 1
    5     1  3        5                          5           4           2
```

If we make a grouping of columns to correspond with the probable interruption key, as shown in figure 46, we note that we have 2 extra groups of 5 and no group of 6 and 7, respectively, as required by a key of the determined length. If, however, we shift the sequence so as to bring columns 27 and 28 to the left, which is perfectly legitimate since we are dealing with a cyclic key, we have:

FIGURE 47

```
27-28-1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26
 2- 1-5-4-3-2-1-1-3-2-1- 5- 4- 3- 2- 1-            5- 4- 3- 2- 1- 4- 3- 2- 1
        7        1  3        5                          5        4
```

We now have our group of 7 columns, but lack the 1 of 6 and 1 of 2; but we have 2 groups of 5 columns. Obviously, the only thing to do is to insert the sequence 2–1–1– in its proper place as follows:

FIGURE 48

```
27-28-1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26
 2- 1-5-4-3-2-1-1-3-2-1- 5- 4- 3- 2- 1- 2- 1- 1- 5- 4- 3- 2- 1- 4- 3- 2- 1
      7       1   3         5        2           6              4
```

We have thus reconstructed the complete interruption key, and may now proceed to transcribe the message in groups in accordance therewith. Since the first group contains 7 letters, and since the letter S, with which we started our first transcription (fig. 34), is labelled 1, it is clear that we must start our second transcription with the two letters preceding S, i.e., the twenty-fifth and twenty-sixth letters of the cipher message. Thus:

FIGURE 49

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

etc., etc.

After this transcription a third set-up is necessary, in which the sequences of letters are reversed within groups, in order to bring the cipher letters back into the original arrangement before transposition. The rearranged set-up is as follows:

FIGURE 50

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

We are now confronted with a rather simple case of the analysis of five *reciprocal* and interrelated alphabets. They are composed of the consolidated frequency tables applying to the columns which belong in the same alphabets and are as follows:

TABLE XXIV

1

Columns 5—6—9—14—16—17—22—26—28

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	10	5	16	27	12	11	2	6	3	17	4	8	25	13	6	6	8		5	9	9	19		1	14

2

Columns 4—8—13—15—21—25—27

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
22	3	4	7	9		5	6		9	19		17	6	12	9	4	3		3	7		15	3	14	4

3

Columns 3—7—12—20—24

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			12	2	12		8	4	4	8		2	1		14	8	5	7	2	2	14	10	8	1	4

4

Columns 2—11—19—23

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	1	2		7	2		13	13		5	11	2	5	9	8	1		1	4	2	3			1	1

5

Columns 1—10—18

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	5		3	9	1		2	5	6	7	8	2	1			3	9			8	3	4		

It will be unnecessary in this paper to discuss the method of deciphering the message by an analysis of these five single-frequency tables. Suffice it to indicate the values obtained from the decipherment. They are given below, where the values obtained from a knowledge of the reciprocal relation are placed within parentheses. Only four values remain unknown, all in alphabet 5.

TABLE XXV

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

We may now attempt a reconstruction of the original, or primary, alphabet, of which these are secondaries. Note the following values:

In alphabet 1, E=N
In alphabet 2, N=O
In alphabet 3, O=V
In alphabet 4, V=M
In alphabet 5, M=T

Now if E occurs in the upper half of alphabet 1, the table of alphabets must contain a column like this:

$$
\begin{array}{rcl}
\text{Alphabet 1}\left\{\begin{array}{c}E\\N\end{array}\right. & & \\
 & \left.\begin{array}{c}N\\O\end{array}\right\} & \text{Alphabet 2} \\
\text{Alphabet 3}\left\{\begin{array}{c}O\\V\end{array}\right. & & \\
 & \left.\begin{array}{c}V\\M\end{array}\right\} & \text{Alphabet 4} \\
\text{Alphabet 5}\left\{\begin{array}{c}M\\T\end{array}\right. & &
\end{array}
$$

Let us assume this to be correct. In alphabet 3, E will again be in the upper half of the alphabet. We have these values:

In alphabet 3, E=D
In alphabet 4, D=Z
In alphabet 5, Z=?

But to this we may add two more values since we have the value of E in alphabet 2. Thus:

In alphabet 1, K=A
In alphabet 2, A=E
In alphabet 3, E=D
In alphabet 4, D=Z
In alphabet 5, Z=?

REF ID:A64722

Since the upper line in alphabet 3 is displaced but one letter to the right as compared with that of alphabet 1, we have two columns in the table as follows:

```
Alphabet 1 { E K
           { N A }
Alphabet 3 { O E } Alphabet 2
           { V D }
Alphabet 5 { M Z } Alphabet 4
           { T
```

We may continue thus:

In alphabet 1, W=I
In alphabet 2, I=K
In alphabet 3, K=N
In alphabet 4, N=H
In alphabet 5, H=C

Hence we have this:

```
Alphabet 1 { E K W
           { N A I }
Alphabet 3 { O E K } Alphabet 2
           { V D N }
Alphabet 5 { M Z H } Alphabet 4
           { ? U C
```

The process is very simple and easy to continue. Finally we have this:

```
Alphabet 1 { E K W F X L Y
           { N A I R Q B J }
Alphabet 3 { O E K W F X L } Alphabet 2
           { V D N A I R Q }
Alphabet 5 { M Z H O E K W } Alphabet 4
           { ? U C S V D N
```

We may now continue from the sequences given already. Thus in the last line we see the sequence . . . S V D N, in the fourth line, V D N A I R Q. Hence, we may add A I R Q to the last line. The same process applied to the other lines gives us:

```
             1 2 3 4 5 6 7 8 9
Alphabet 1 { E K W F X L Y             M Z H O
           { N A I R Z B J   U         C S V D }
Alphabet 3 { O E K W F X L Y             M Z H } Alphabet 2
           { V D N A I R Q B J .         U C S }
Alphabet 5 { M Z H O E K W F X L Y             } Alphabet 4
           { ? U C S V D N A I R Q B J
```

We may fill in the rest from the alphabets themselves, and we have:

FIGURE 51

	1	2	3	4	5	6	7	8	9	10	11	12	13	
Alphabet 1	E	K	W	F	X	L	Y	G	P	M	Z	H	O	
	N	A	I	R	Q	B	J	T	U	C	S	V	D	Alphabet 2
Alphabet 3	O	E	K	W	F	X	L	Y	G	P	M	Z	H	
	V	D	N	A	I	R	Q	B	J	T	U	C	S	Alphabet 4
Alphabet 5	M	Z	H	O	E	K	W	F	X	L	Y	G	P	
	T	U	C	S	V	D	N	A	I	R	Q	B	J	

Taking alphabet 1, a speedy reconstruction of the original rectangle is at once effected. Thus:

FIGURE 52

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

The keyword is EXPORTS. In conformity with the agreements of the system, the indicators should be K X Y and indicate the following:

K, the initial letter of the distorted alphabet;
X, the fifth letter in alphabet 1, hence five alphabets;
Y, the seventh letter in alphabet 1, hence seven groups arranged as follows:

```
E K W F X L Y
1 3 5 2 6 4 7
```

The indicators will be after either the fifth letter or the nineteenth (corresponding to the numerical value of the initial or final letter of the keyword). We find K X Y after the nineteenth letter.

We may now proceed to decipher the first few groups of the message and the entire solution is at hand. It is as follows:

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

               1   3       5
               1   1 2 3   1 2 3 4 5
               P   D K U   L I S Y Z  . . . . .
               P   U K D   Z Y S I L  . . . . .
               U   P I E   S T H E R  . . . . .
```

"A hostile reinforced brigade occupies the . . ."

2. **Solution of a series of messages.**—The solution of a series of messages in this cipher presents an interesting demonstration of the fact that in cryptography it is often the more insignificant details of a system that enable the cryptanalyst to solve the enemy's messages rather than any definite weakness of the method from the cryptographic point of view. In this case, the solution of a series of messages based upon the same keyword, but involving all the manifold modifications of alphabets and interruption key, can be achieved without attempting the decipherment of a single message. The method involves merely an analysis of the indicators for a series of messages, resulting in a direct and speedy reconstruction of the various generating rectangles derived from the same keyword.

In the first place, it may be pointed out at once that the various generating rectangles based upon the same keyword consist of two parts: A constant sequence, consisting of the *keyword proper*, making up the first line of the rectangle, and a variable, or revolving sequence, consisting of the remaining letters of the alphabet, or as we shall term it, the *"residual sequence"*, making up the remaining lines of the rectangle.

In the second place, the length of the interruption key for each message can be determined by applying the principles of coincidence as explained on page 68. Each message is then accompanied by the number thus found.

Given the beginnings of a series of 18 messages, the lengths of whose interruption keys have been determined and are as indicated below, let us proceed to an analysis of these lengths, which in a short time will lead to a direct reconstruction of the keyword.

FIGURE 53.—*Beginnings of series of messages*

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

Let us now turn our attention to the generating rectangles that are possible, indicating merely their outlines, for keywords of 5 to 10 letters in length, and numbering the squares which will contain the first, second, third, to the ninth letters of the first line of the alphabet table. We will assume that no keyword will consist of more than 10 letters, and no interruption key of more than 9 numbers. The principles to be elucidated may be extended by the reader to cover other cases.

FIGURE 54

(A)

1	2	3	4	5
1	7			
2	8			
3	9			
4				
5				
6				

(B)

1	2	3	4	5	6
1	6				
2	7				
3	8				
4	9				
5					

(C)

1	2	3	4	5	6	7
1	5	9				
2	6					
3	7					
4	8					

(D)

1	2	3	4	5	6	7	8
1	5	9					
2	6						
3	7						
4	8						

(E)

1	2	3	4	5	6	7	8	9
1	4	7						
2	5	8						
3	6	9						

(F)

1	2	3	4	5	6	7	8	9	10
1	4	7							
2	5	8							
3	6	9							

Note now that in (A) of figure 54, the seventh letter of line 1 of the alphabet table based upon a keyword of 5 letters will be the second letter of the keyword proper and, as such, does not change its position. This follows from the method of constructing the alphabets from the generating rectangle. In other words, with a 5-letter keyword, no matter what letters be chosen as the initial letters of the various possible generating rectangles, all numerical keys consisting of seven numbers will always be designated by the same letter in the indicator group, because the first line of all generating rectangles based upon the same keyword is always the same. Since the letter which designates the length of the interruption key is the third letter of the indicator group for a keyword of 5 letters, all messages which factor for a key of seven numbers must show the same letter as the third element of the indicator group. Conversely, if messages which factor for a key of seven numbers show a constancy with respect to the third element of the indicator group, then it must follow that the generating rectangle is based upon a 5-letter keyword, and that the corresponding indicator letter is the second letter of the keyword proper. Similarly, when the third element of the indicator group is constant in messages which factor for a key of six numbers, a generating rectangle based upon a 6-letter keyword is indicated and the corresponding indicator letter is the second letter of the keyword proper. When the third elements of *two sets* of indicator groups are constant and the interruption keys for the corresponding sets of messages consist of five and nine numbers, a 7- or 8-letter keyword is indicated and the corresponding indicator letters are the second and third letters of the key-

TABLE XXVI

Message number	Indicator group	Length of interruption key	Message number	Indicator group	Length of interruption key
6	B S F	8	3	J X U	4
11	D S F	5	15	K U O	5
14	D U F	5	2	K Y O	5
8	F M H	8	7	O B Q	8
18	F U U	4	13	O Z U	4
5	G W M	7	1	O P B	6
4	H M K	8	16	O P B	6
17	H W U	4	12	P D S	8
9	J Y O	8	10	Q D M	7

From the arrangement of the messages in figure 55 and this list we conclude the following:

(1) The keyword consists of either 9 or 10 letters.

(2) The letter L is either the initial or final letter of the keyword proper.

(3) The letters UM form the second and third letters of the keyword proper.

(4) The letters B, D, F, G, H, J, K, O, P, Q are not in the keyword proper.

We now take each indicator group and from its accompanying data make certain deductions with respect to the sequence of letters in the keyword alphabet. Thus, for example, the indicator group B S F, applying to a message, the length of whose interruption key is 8, means that certain of the letters within the generating rectangle concerned are as shown in figure 56, where both possibilities as regards the length of the keyword are indicated.

FIGURE 56

(A)

1	2	3	4	5	6	7	8	9
	U	M						
B		F						

(B)

1	2	3	4	5	6	7	8	9	10
	U	M							
B		F							

Since there is room for but one letter between B and F, it follows that only one of the intervening letters, C, D, and E, is left in the residual sequence, the other two being in the keyword proper. Now the indicator group D S F shows that D is the initial letter for the generating rectangle for message 11, and, since a letter which can be an initial letter of an indicator group cannot be a letter of the keyword proper, it follows that the order in the residual sequence is B D F, and that the letters C and E are in the keyword proper.

This is confirmed by the indicator group D S F, which accompanying message 11 has an interruption key consisting of five numbers. Thus:

FIGURE 57

(A)

1	2	3	4	5	6	7	8	9
	U	M						
D	F							

(B)

1	2	3	4	5	6	7	8	9	10
	U	M							
D	F								

Again, the indicator group F M H, applying to message 8, whose interruption key consists of eight numbers, shows that the sequence must be B D F G H. Thus:

FIGURE 58

(A)

1	2	3	4	5	6	7	8	9
	U	M						
F		H						

(B)

1	2	3	4	5	6	7	8	9	10
	U	M							
F		H							

There is one space vacant between F and H, which must be occupied by G.

The conclusions furnished by the other indicator groups are given in the diagrams below. The question as to whether a 9-letter or a 10-letter keyword is involved will be determined later.

FIGURE 59

	1	2	3
		U	M
HMK—8	H		K

H (I J) K

Since J is present as an initial letter the sequence must be: B D F G H J K

	1	2	3
		U	M
JYO—8	J		O

J (K L M) O

Since K is present as an initial letter, the sequence must be: B D F G H J K O

	1	2	3
		U	M
OBQ—8	O		Q

O ( ) Q

Since P falls naturally between O and Q, the sequence must be: B D F G H J K O P Q

In the absence of additional indicator groups, with different initial letters, we cannot continue in the same manner and reconstruct the entire sequence. Let us, however, try to determine now whether the keyword contains 9 or 10 letters. With the residual sequence as determined above, let us fill in the generating rectangle for the indicator group D S F, according to both assumed keyword lengths. Thus:

FIGURE 60

(A)

1	2	3	4	5	6	7	8	9
	U	M						
D	F	G	H	J	K	O	P	Q
S							B	

(B)

1	2	3	4	5	6	7	8	9	10
	U	M							
D	F	G	H	J	K	O	P	Q	S

Now the interruption key for the indicator group D S F consists of five numbers. The system is such that the number of alphabets employed in a message must be equal to or less than the length of the interruption key; it cannot be more than this length. Hence, the letter S of the indicator group D S F must indicate a number of alphabets equal to five, or less than five. Upon the basis of a 10-letter keyword, the number of alphabets indicated by the letter S would be impossible; but upon the basis of a 9-letter keyword, the number of alphabets would be three, which is very probable. We may tentatively consider it as established that the keyword is 9 letters in length.

Note now diagram (A) of figure 60. There are six spaces vacant at the end of the keyword alphabet. The letters from S to B in the normal alphabet are T U V W X Y Z A, a total of eight. Now we know that U is in the keyword proper so that the residual sequence does not contain this letter. Of the letters, V W X Y Z A, the one most likely to be present in the keyword proper is A, leaving the sequence V W X Y Z as the end of the key-word alphabet.

The keyword proper consists of the letters not present in the residual sequence. It must, therefore, consist of the letters A, C, E, I, L, M, N, R, and U. We know that the letters U M form the second and third letters of the word; and the position of the indicator groups in the cipher text makes L very probable as the initial or final letter of the word. Now, very few words beginning with L U M and containing the other letters A, C, E, I, N, and R can be found. But if we assume L to be the final letter of the keyword, the most probable ending would be C A L. Given . U M . . . C A L , the word NUMERICAL soon suggests itself.

With the keyword at hand, every generating rectangle can be constructed at once, and the messages may now be deciphered as rapidly as by the legitimate recipient.

The decipherment of message 1 follows herewith:

FIGURE 61

K I B I Y F P P L H G Z O P B M A W F V T B N B I F U E B . . .

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

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

"Zero hour will be at two o'clock . . . "

The fatal defect in this system lies in the fact that a key is used which introduces a frequently repeated cycle within a message. The determination of the length of this cycle, and its reconstruction by means of a comparison of alphabets based upon the index of coincidence, enables a speedy solution to be attained. The insertion of the key indicators within messages makes the reconstruction of the keyword and the consequent solution of a series of messages very easy. The many details involved in encipherment, and decipherment, concomitants of an attempt to make the entire operations dependent upon the knowledge of a single keyword and the ease with which a solution may be achieved in the case of a single message or a series of messages makes this cipher unsafe for use in either the field or the more important operations of the larger headquarters in the rear.

O