

# The Applications of Probability to Cryptography

Alan M. Turing

arXiv:1505.04714v1 [math.HO] 18 May 2015

## Copyright

The underlying manuscript is held by the National Archives in the UK and can be accessed at [www.nationalarchives.gov.uk](http://www.nationalarchives.gov.uk) using reference number HW 25/37. Readers are encouraged to obtain a copy.

The original work was under Crown copyright, which has now expired, and the work is now in the public domain.

## Editor's Notes

### Provenance

Two Second World War research papers by Alan Turing were declassified recently. The papers, “*The Applications of Probability to Cryptography*” and its shorter companion “*Statistics of Repetitions*”, are available from the National Archives in the UK at [www.nationalarchives.gov.uk](http://www.nationalarchives.gov.uk).

The released papers give the full text, along with figures and tables, and provide a fascinating insight into the preparation of the manuscripts, as well as the style of writing at a time when typographical errors were corrected by hand, and mathematical expression handwritten into spaces left in the text.

Working with the papers in their original format provides some challenges, so they have been typeset for easier reading and access. We recommend that the typeset versions are read with a copy of the original manuscript at hand.

This document contains the text and figures for “*The Applications of Probability to Cryptography*”, the companion paper is also available in typeset form from arXiv at [www.arxiv.org](http://www.arxiv.org). These notes apply to both documents.

Separately, a journal article by Zabell<sup>1</sup> provides an analysis of the papers and further background information.

### The text

It is not our intent to cast Alan Turing's manuscripts into a journal style article, but more to provide clearer access to his writing and, perhaps, to answer the questions “If Turing had have had access to typesetting software, what would his paper have looked like?”. Consequently no “house-style” copy-editing has been imposed. Occasional punctuation has been added to improve readability, some obvious errors corrected, and paragraph breaks added to ease the reading of long text blocks - and occasionally to give a better text flow. Turing uses typewriter underlining, single, and double quotes to indicate emphasis or style; these have been implemented using font format changes, double quotes are used as needed.

The manuscript has many typographical errors, deletions, and substitutions, all of which are indicated by over-typing, crossed out items, and handwritten pencil or ink annotations. These corrections have been implemented in this document to give the text that we presume Turing intended. Additionally, there are some hand written notes in the manuscript, which may or may not be by Turing; these are indicated by the use of footnotes.

British English spelling is used in the manuscript and this is retained, so words such as favour, neighbourhood, cancelling, etc. will be encountered. Turing appears to favor the spellings bigramme, trigramme, tetragramme, etc., although he is not always consistent; throughout this document the favored rendering is used.

Turing's wording is unchanged to give the full flavor of his original style. This means that “That is to say that we suppose for instance that ..... ” will be encountered - amongst others!

Both papers end abruptly, no summary or conclusion is offered, perhaps the papers are incomplete or pages are missing. To indicate the end of the manuscript we have marked the end of each paper with a printing sign - an infinity symbol between two horizontal bars.

---

<sup>1</sup>Zabell, S. 2012. “Commentary on Alan M.Turing: The Applications of Probability to Cryptography” *Cryptologia*, 36:191-214.

In the section on a letter subtractor problem, reference is made to other methods to be discussed later in the paper. Thus does not happen - perhaps another indicator of an incomplete paper or missing pages.

Finally, Turing uses some forward page references that appear in the manuscript as *see(p)*, obviously intending to return and complete the reference. This does not happen, so these references remain unresolved.

In short, we strive to represent Turing's text as he wrote it.

### **Ciphertext, cleartext, etc.**

In an attempt to capture the flavor of the time, ciphertext, cleartext, keys, etc. are displayed in a fixed pitch, bold, non-serif font to represent the typewriter, teletype, and telegraph machines that would have printed the original code, *viz.* **CONDITIONS**.

### **Mathematics**

In the manuscript all mathematics is hand written in ink and pencil in spaces left between the typed text. Sometimes adequate space was left, other time not, and the handwriting spills into margins and adjacent lines, adding to the reading challenge. We have cast all mathematics into standard in-line or display formats as appropriate. We have used the `mathcal` font in places to capture the flavor of Turing's handwriting, *e.g.* "the probability  $p$ " appears as "the probability  $\mathcal{P}$ ".

Turing uses no punctuation in his mathematics, this has been added to be consistent with modern practice<sup>2</sup>; he also uses letters to reference equations - numbers are used in this document. In many places we have added parentheses to give clarity to an expression, and in some places where Turing is inconsistent in his uses of parentheses for a mathematical phrase (the expression for letter probability in the Vigenère in particular) we have chosen one format and been consistent in its use.

As Turing demonstrates a love of dense mathematics the algebraic multiplication symbol  $\times$  has occasionally been used for readability, so all standard forms of multiplication will be encountered, *viz.*,  $ab$ ,  $a \times b$ ,  $a \cdot b$ . Finally, convention suggests that the subject of a formula or expression sits on its own on the left hand side of the equals sign, with the subsidiary variables collected on the right hand side. Turing adheres to this convention as it suits him, his preference is retained.

In short, we strive to retain the elegance of Turing's mathematics, whilst casting it into a modern format.

### **Figures and tables**

All figures have been included with rearrangement of some items to improve clarity or document flow. Turing uses a variety of papers, styles, inks, pen, and pencil; these have all been represented in standard figure and table format.

### **Contents page**

Turning provides a rudimentary Contents for "*The Applications of Probability to Cryptography*", this has been reworked with some additions to make it more meaningful. "*Statistics of Repetitions*", being much shorter, requires no Contents.

### **Editors**

The editor can be contacted at: [ian.taylor@maths.oxon.org](mailto:ian.taylor@maths.oxon.org).

---

<sup>2</sup>See, for instance, Higham, Nicholas J. 1998. *Handbook of writing for the Mathematical Sciences*, SIAM, Philadelphia.

## Contents

Chapter 1. Introduction	1
1.1. Preamble	1
1.2. Meaning of probability and odds	1
1.3. Probabilities based on part of the evidence	1
1.4. <i>A priori</i> probabilities	1
1.5. The Factor Principle	3
1.6. Decibanage	4
Chapter 2. Straightforward Cryptographic Problems	5
2.1. Vigenère	5
2.2. A letter subtractor problem	11
2.3. Theory of repeats	14
2.3.1. Simple form of theory	15
2.3.2. Second simplified form of theory	16
2.3.3. General form of theory	16
2.4. Transposition ciphers	20
2.4.1. A probability problem	20
2.4.2. The Probability Formula	24
2.4.3. Another probability problem	26



## CHAPTER 1

# Introduction

### 1.1. Preamble

The theory of probability may be used in cryptography with most effect when the type of cipher used is already fully understood, and it only remains to find the actual keys. It is of rather less value when one is trying to diagnose the type of cipher, but if definite rival theories about the type of cipher are suggested it may be used to decide between them.

### 1.2. Meaning of probability and odds

I shall not attempt to give a systematic account of the theory of probability, but it may be worth while to define shortly *probability* and *odds*. The *probability* of an event on certain evidence is the proportion of cases in which that event may be expected to happen given that evidence. For instance if it is known the 20% of men live to the age of 70, then knowing of Hitler only *Hitler is a man* we can say that the probability of Hitler living to the age of 70 is 0.2. Suppose that we know that *Hitler is now of age 52* the probability will be quite different, say 0.5, because 50% of men of 52 live to 70.

The *odds* of an event happening is the ratio  $\mathcal{P}/(1-\mathcal{P})$  where  $\mathcal{P}$  is the probability of it happening. This terminology is connected with the common phraseology *odds of 5:2 on* meaning in our terminology that the odds are 5/2.

### 1.3. Probabilities based on part of the evidence

When the whole evidence about some event is taken into account it may be extremely difficult to estimate the probability of the event, even very approximately, and it may be better to form an estimate based on a part of the evidence, so that the probability may be more easily calculated. This happens in cryptography in a very obvious way. The whole evidence when we are trying to solve a cipher is the complete traffic, and the events in question are the different possible keys, and functions of the keys. Unless the traffic is very small indeed the theoretical answer to the problem “What are the probabilities of the various keys?” will be of the form “The key . . . has a probability differing almost imperceptibly from 1 (certainty) and the other keys are virtually impossible”. But a direct attempt to determine these probabilities would obviously not be a practical method.

### 1.4. *A priori* probabilities

The evidence concerning the possibility of an event occurring usually divides into a part about which statistics are available, or some mathematical method can be applied, and a less definite part about which one can only use one’s judgement.

Suppose for example that a new kind of traffic has turned up and that only three messages are available. Each message has the letter V in the 17th place and G in the 18th place. We want to know the probability that it is a general rule that we should find V and G in these places. We first have to decide how probable it is that a cipher would have such a rule, and as regards this one can probably only guess, and my guess would be about 1/5,000,000. This judgement is not entirely a guess; some rather insecure mathematical reasoning has gone into it, something like this:-

The chance of there being a rule that two consecutive letters somewhere after the 10th should have certain fixed values seems to be about 1/500 (this is a complete guess). The chance of the letters being the 17th and 18th is about 1/15 (another guess, but not quite as much in the air). The probability of a letter being V or G is 1/676 (hardly a guess at all, but expressing a judgement that there is no special virtue in the bigramme VG). Hence the chance is 1/(500 × 15 × 676) or about 1/5,000,000. This is however all so vague, that it is more usual to make the judgment “1/5,000,000” without explanation.

The question as to what is the chance of having a rule of this kind might of course be resolved by statistics of some kind, but there is no point in having this very accurate, and of course the experience of the cryptographer itself forms a kind of statistics.

The remainder of the problem is then solved quite mathematically. Let us consider a large number of ciphers *chosen at random*.  $N$  of them say. Of these  $N/5,000,000$  of them will have the rule in question, and the remainder not. Now if we had three messages of each of the ciphers before us, we should find that for each of the ciphers with the rule, three messages have VG in the required place, but of the remaining  $(4,999,999 \times N)/5,000,000$  only a proportion  $1/676^3$  will have them. Rejecting the ciphers which have not the required characteristics we are left with  $N/5,000,000$  cases where the rule holds, and  $(4,999,999 \times N)/(5,000,000 \times 676^3)$  cases where it does not. This selection of ciphers is a random selection of ones which have all the known characteristics of the one in question, and therefore the odds in favour of the rule holding are:

$$\frac{N}{5,000,000} : \frac{4,999,999 \times N}{5,000,000 \times 676^3},$$

*i.e.*  $676^3 : 4,999,999,$   
or about 60 : 1 on.

It should be noticed that the whole argument is to some extent fallacious, as it is assumed that there are only two possibilities, *viz.* that either VG must always occur in that position, or else that the letters in the 17th and 18th positions are wholly random. There are however many other possibilities worth consideration, *e.g.*

On the day in question we have VG in the position in question.

Or on another day we have some other fixed pair of letters.

Or in the positions 17, 18 we have to have one of the four combinations VG, RH, OM, IL and by chance VG has been chosen for all the three messages we have had.

Or the cipher is a simple substitution and VG is the substitute of some common bigramme, say TH.

The possibilities are of course endless, and it is therefore always necessary to bear in mind the possibility of there being other theories not yet suggested.



The *a priori* probability sometimes has to be estimated as above by some sort of guesswork, but often the situation is more satisfactory. Suppose for example that we know that a certain cipher is a simple substitution, the keys having no specially noticeable properties. Suppose also that we have 50 letters of such a message including five occurrences of P. We want to know how probable it is that P is the substitute of E. As before we have to answer two questions.

- (1) How likely is it that P would be the substitute of E neglecting the evidence of the five Es occurring in the message?
- (2) How likely are we to get 5 Ps?
  - (a) If P is not the substitute of E
  - (b) If P is the substitute of E.

I will not attempt to answer the second question for the present. The answer to the first is simply that the probability of a letter being the substitute of E is independent of what the letter is, and is therefore always  $1/26$ , in particular it is  $1/26$  for the letter P. The only guesswork here is the judgement that the keys are chosen at random.

### 1.5. The Factor Principle

Nearly all applications of probability to cryptography depend on the *factor principle* (or Bayes' Theorem). This principle may first be illustrated by a simple example. Suppose that one man in five dies of heart failure, and that of the men who die of heart failure two in three die in their beds, but of men who die from other causes only one in four die in their bed, but of the men who die from other causes only one in four dies in their beds. (My facts are no doubt hopelessly inaccurate). Now suppose we know that a certain man died in his bed. What is the probability that he died of heart failure? Of all numbering  $N$  say we find that

$N \times (1/5) \times (2/3)$	<i>die in their beds of heart failure</i>
$N \times (1/5) \times (1/3)$	... <i>elsewhere</i> .....
$N \times (4/5) \times (1/4)$	<i>die in their beds from other causes</i>
$N \times (4/5) \times (3/4)$	... <i>elsewhere</i> .....

Now as our man died in his bed we do not need to consider the cases of men who did not die in their beds, and these consist of

$N \times (1/5) \times (2/3)$	<i>cases of heart failure and</i>
$N \times (4/5) \times (1/4)$	<i>from other causes</i>

and therefore the odds are  $1 \times (2/3) : 4 \times (1/4)$  in favour of heart failure. If this had been done algebraically the result would have been

$$\begin{aligned}
 & \text{A posteriori odds of the theory} \\
 & = \text{A priori odds of the theory} \\
 & \quad \times \frac{\text{Probability of the data being fulfilled if the theory is true}}{\text{Probability of the data being fulfilled if the theory is false}}.
 \end{aligned}$$

In this the *theory* is that the man died of heart failure, and the *data* is that he died in his bed. The general formula above will be described as the *factor principle*, the

ratio

$$\frac{\text{Probability of the data if the theory is true}}{\text{Probability of the data if the theory is false}}$$

is called the factor for the theory on account of the data.

### 1.6. Decibanage

Usually when we are estimating the probability of a theory there will be several independent pieces of evidence *e.g.* following our last example, where we want to know whether a certain man died of heart failure or not, we may know

- (1) He died in his bed
- (2) His father died of heart failure
- (3) His bedroom was on the ground floor

and also have statistics telling us

- (a) 2/3 of men who die of heart failure die in their beds
- (b) 2/5 ..... have fathers who died of heart failure
- (c) 1/2 ..... have bedroom on the ground floor
- (d) 1/4 of men who died from other causes die in their beds
- (e) 1/6 ..... have fathers who died of heart failure
- (f) 1/20 of men who die of other cause have their bedrooms on the ground floor

Let us suppose that the three pieces of evidence are independent of one another if we know that he died of heart failure, and also if we know that he did not die of heart failure. That is to say that we suppose for instance that knowing that he slept on the ground floor does not make it any more likely that he died in his bed if we knew all along that he died of heart failure. When we make these assumptions the probability of a man who died of heart failure satisfying all three conditions is obtained simply by multiplication, and is  $(2/3) \times (2/5) \times (1/2)$  and likewise for those who died from other causes the probability is  $(1/4) \times (1/6) \times (1/20)$ , and the factor in favour of the heart theory failure is

$$\frac{(2/3) \times (2/5) \times (1/2)}{(1/4) \times (1/6) \times (1/20)}$$

We may regard this as the product of three factors  $(2/3)/(1/4)$  and  $(2/5)/(1/6)$  and  $(1/2)/(1/20)$  arising from from the three independent pieces of evidence. Products like this arise very frequently, and sometimes one will get products involving thousands of factors, and large groups of these factors may be equal. We naturally therefore work in terms of the logarithms of the factors. The logarithm of the factor, taken to the base  $10^{1/10}$  is called *decibanage in favour of the theory*. A *deciban* is a unit of evidence; a piece of evidence is worth a deciban if it increase the odds of the theory in the ratio  $10^{1/10} : 1$ . The deciban is used as a more convenient unit than the *ban*. The terminology was introduced in honor of the famous town of Banbury.

Using this terminology we might say that the fact that our man died in bed scores 4.3 decibans in favour of the heart failure theory ( $10 \log(8/3) = 4/3$ ). We score a further 3.8 decibans for his father dying of heart failure, and 10 for his having his bedroom on the ground floor, totalling 18.1 decibans. We then bring in the *a priori* odds 1/4 or  $10^{-6/10}$  and the result is the the odds are  $10^{12.1/10}$ , or as we may say "12.1 deciban up on evens". This means about 16:1 on.

## Straightforward Cryptographic Problems

### 2.1. Vigenère

The factor principle can be applied to the solutions of a Vigenère problem with great effect. I will assume here that the period of the cipher has already been determined. Probability theory may be applied to this part of the problem also, but that is not so elementary. Suppose our cipher, written out in its correct period is<sup>1</sup>

```

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

```

FIGURE 1. Vigenère problem.

*(It is only by chance that it makes a rectangular array.)*

Let us try to find the key for the first column, and for the moment let us only take into account the evidence afforded by the first letter D. Let us first consider the key B. The factor principle tells us

$$\begin{aligned} \text{Odds in favour of key B} &= \text{A priori odds in favour of key B} \\ &\times \frac{\text{Probability of getting D in cipher if key is B}}{\text{Probability of getting key D in cipher if key is not B}} \end{aligned}$$

Now the *a priori* odds in favour of key B may be taken as 1/25. The probability of getting D in the cipher with the key B is just the probability of getting C in the clear which (using the count on 1000 letters in Fig 2) is 0.021. If however the key is not B we can have any letter other the C in the clear, and the probability is (1 - 0.021)/25. Using the evidence of the D then the odds in favour of the key B are

$$\frac{1}{25} \times \left( \frac{25 \times 0.021}{1 - 0.021} \right).$$

---

<sup>1</sup> Turing reverses the N M in the first line, which should read DKQHSZMNP. The right sequence is used in Fig 5, which gives the correct decode.

We may then consider the effect of the next letter in the column R which gives a further factor of  $(25 \times 0.064)/(1 - 0.064)$ . We are here assuming that the evidence of the R is independent of the evidence of the D. This is not quite correct, but is a useful approximation; a more accurate method of calculation will be given later. Let us write  $\mathcal{P}_\alpha$  for the frequency of the letter  $\alpha$  in plain language. Then our final estimate for the odds in favour of key B is

$$\frac{1}{25} \prod_i \frac{25\mathcal{P}_{\alpha_i-1}}{1 - \mathcal{P}_{\alpha_i-1}}.$$

where  $\alpha_1, \alpha_2, \dots$  is the series of letters in the 1st column, and we use the letters and numbers interchangeably, A meaning 1, B meaning 2, ..., Z meaning 26 or 0. More generally for key  $\beta$  the odds are

$$\frac{1}{25} \prod_i \frac{25\mathcal{P}_{\alpha_i-\beta+1}}{1 - \mathcal{P}_{\alpha_i-\beta+1}}.$$

The value of this can be calculated by having a table of the decibanage corresponding to the factors  $25\mathcal{P}_\alpha/(1 - \mathcal{P}_\alpha)$ . One then decodes the column with the various possible keys, looks up the decibanage, and adds them up.

The most convenient form for doing this is a table of values of  $20 \log_{10}[25\mathcal{P}_\alpha/(1 - \mathcal{P}_\alpha)]$ , taken to the nearest integer, or as we may say, the values of the score in *half decibans*. One may also have columns showing multiples of these, and the table made of double height<sup>2</sup> (Fig 3). For the first column with key B the decoded column is CQWS●●OAV,<sup>3</sup> and we score -5 for C, -26 for Q, -5 for W, 17 for the three letters S, 5 for O, 7 for A and -10 V, totalling -17. These calculations can be done very quickly by the use of the transparent gadget Fig 4, in which squares are ringed in pencil to show the number of letters occurring in the column.

A	84	J	2	S	73
B	23	K	5	T	81
C	21	L	38	U	19
D	46	M	34	V	11
E	116	N	66	W	21
F	20	O	66	X	16
G	25	P	15	Y	24
H	49	Q	2	Z	3
I	76	R	64		

FIGURE 2. Count on 1000 letters.

(English text)

*The value for X has been taken more or less at random as a compromise between real language & telegraphese. Also I added to each entry (see p)<sup>4</sup>.*

<sup>2</sup> Turing provides a table of double height for Fig 3 to allow the “gadget” of Figure 4 to be used with any letter of the alphabet as a decode key - hence the double alphabet. Figure 4 can be prepared as a transparency, with the original markings cleared, and markings for the new decode letter added. Fig 3 and Fig 4 are correctly proportioned in this document for this to work.

<sup>3</sup> S●● means SSS, for a total of three letter S, as noted in the following arithmetic. The linear decode for the example is CQWSSOAVS

<sup>4</sup> Forward reference left unresolved in the manuscript.

31	26	20	13	7	A
-23	-18	-14	-9	-5	B
-26	-21	-16	-10	-5	C
7	6	4	3	1	D
48	38	29	19	10	E
-28	-22	-17	-11	-6	F
-19	-15	-11	-8	-4	G
10	8	6	4	2	H
29	23	17	12	6	I
-131	-103	-77	-52	-26	J
-99	-79	-59	-40	-20	K
-2	-2	-1	-1	0	L
-6	-5	-4	-2	-1	M
23	18	14	9	5	N
23	18	14	9	5	O
-41	-33	-25	-16	-8	P
-131	-103	-77	-52	-26	Q
22	18	13	9	4	R
28	22	17	11	6	S
32	26	19	13	6	T
-31	-25	-19	-12	-6	U
-54	-43	-32	-22	-10	V
-26	-21	-16	-10	-5	W
-38	-30	-23	-15	-8	X
-20	-16	-12	-8	-4	Y
-111	-89	-67	-44	-22	Z
31	26	20	13	7	A
-23	-18	-14	-9	-5	B
-26	-21	-16	-10	-5	C
7	6	4	3	1	D
48	38	29	19	10	E
-28	-22	-17	-11	-6	F
-19	-15	-11	-8	-4	G
10	8	6	4	2	H
29	23	17	12	6	I
-131	-103	-77	-52	-26	J
-99	-79	-59	-40	-20	K
-2	-2	-1	-1	0	L
-6	-5	-4	-2	-1	M
23	18	14	9	5	N
23	18	14	9	5	O
-41	-33	-25	-16	-8	P
-131	-103	-77	-52	-26	Q
22	18	13	9	4	R
28	22	17	11	6	S
32	26	19	13	6	T
-31	-25	-19	-12	-6	U
-54	-43	-32	-22	-10	V
-26	-21	-16	-10	-5	W
-38	-30	-23	-15	-8	X
-20	-16	-12	-8	-4	Y
-111	-89	-67	-44	-22	Z

FIGURE 3. Table for scoring a Vigenère.  
In units of half a deciban.

September 13th **B**

B	2.0	3.7	3.1	0.3	0.3	3.7	0.2	1.4	2.0	0.3	2.0	0.3	0.3	0.3	0.3	0.3	0.3	0.3	1.4	0.3	2.0	1.4	3.7	O	144/37	7
B	3.7	2.0	3.7	4.8	3.1	5.4	1.4	3.7	0.3	5.4	0.3	2.0	4.8	3.0	2.0	2.0	0.3	0.3	3.7	4.8	3.7	0.3	3.1	J	100-75	7
B	1.6	6.0	1.6	5.2	1.9	0.8	2.5	3.3	0.8	5.3	3.3	5.0	4.3	5.0	4.3	0.9	1.9	5.0	0.8	4.8	5.0	0.8	1.4	K	112-87	4
B	8.7	4.0	2.7	5.7	2.6	0.8	2.6	0.2	5.2	2.2	1.2	1.2	9.1	1.1	0.2	1.0	3.6	0.3	6.4	2.3	0.1	0.1	0.7	F	204-179	4
B	6.0	6.2	4.9	2.6	6.0	3.6	4.7	2.2	2.6	1.4	1.8	3.8	2.3	2.1	4.1	3.4	2.7	0.1	3.8	7.1	3.7	6.0	0.8	R	204-179	4
B	3.9	3.7	4.0	2.7	4.1	2.4	3.7	3.7	5.6	0.9	4.1	2.5	2.6	2.3	1.3	0.8	2.4	0.6	1.0	4.1	4.6	1.4	1.8	V	172-147	6
B	0.5	1.2	3.0	5.2	3.5	0.1	3.9	2.2	3.9	5.3	0.6	1.5	3.9	2.2	0.5	5.2	2.2	1.8	3.5	3.2	3.5	1.9	2.9	M	120-95	4
B	3.7	2.0	2.0	1.4	2.0	0.0	5.4	3.7	3.7	2.0	3.7	5.4	3.7	3.1	0.3	1.4	9.9	5.4	0.3	2.0	5.4	3.7	4.8	T	88-63	7

*Editor - This table is not referenced in the manuscript, does not have a page number, and nothing appear to indicate its purpose. It is presented here for completeness.*

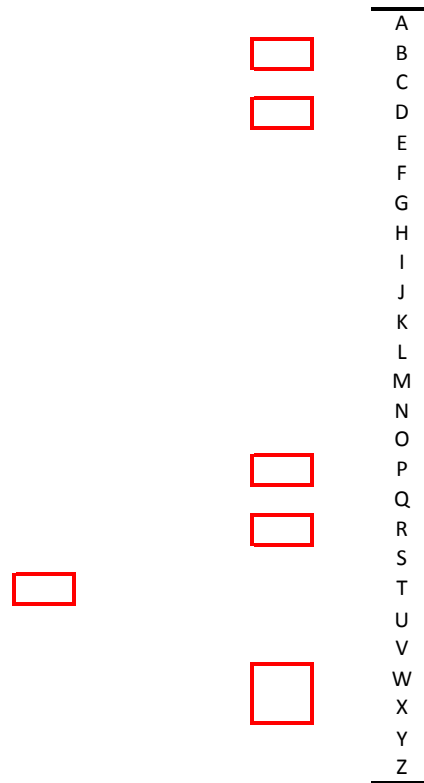


FIGURE 4. Apparatus for scoring a Vigenère.  
Pencil marks arranged for 1st wheel of Fig. 1.

The gadget may be placed over Fig 3 in various positions corresponding to the various keys. The score is obtained by adding up the numbers showing through the various squares. In Fig 5 the alphabet has been written in a vertical below the cipher text of Fig 1, each letter representing a possible key. The score for each key has been written opposite the key, and under the relevant column. An X denotes a bad score, not worth adding up. Usually these will be -15 or worse. It will be seen that for the first column P, having a score of 43 is extremely likely to be right, especially as there is no other score better than 8. If we neglect this latter fact the odds for the key are  $(1/25)10^{2.15}$  *i.e.* about 5:1 on. The effect of decoding this column with key P has been shown underneath.

For the second column the best key is O, but is by no means so certain as the first column. The decode for this column is also shown, and provides very satisfactory combinations with the first column, confirming both keys. (This confirmation could also be based on probability theory, given a table of bigramme frequencies). In the third column I and C are best although D would be very possible, and in the fourth column Q and U are best.

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

<i>Scores for possible keys</i>	A	-2	X	•	X	X		
	B	•	-17	X	X	•	-2	
	C	X	•	3	16	••	X	
	D	•	X	9	9	X		
	E	-6	X	X	X			
	F	X	X	X	X			
	G	X	X	X	-3			
	H	X	••	1	-3	•	X	
	I	X	X	17	X			
	J	X	X	X	•	13		
	K	X	•	X	••	X	-15	
	L	2	X	X	X			
	M	X	X	X	X			
	N	X	X	X	X			
	O	X	•	28	X	•	X	
	P	•	43	X	•	X	X	
	Q	X	X	•	X	22		
	R	•	X	X	X	X		
	S	X	X	-6	X			
	T	•••	8	X	-15	X		
	U	X	•	X	•	X	•	22
	V	X	X	•	X	X		
	W	•	X	••	16	•	1	X
	X	•	X	X	-15	•	X	
	Y	X	X	-18	•	X		
	Z	X	•	-13	•	X	X	
<i>Best Keys</i>		P	O	IC	QU			
		O	W	IO	RN	G		
		C	O	NT	HD	I		
		I	T	HN	EA	S		
<i>Possible Decodes</i>		E	I	MW	TF	O		
		E	T	OU	MI	M		
		A	L	CI	YU	L		
		M	A	CI	LH	I		
		H	I	SY	MI	S		
		E	G	RX	IE	T		

FIGURE 5. Scoring and solving a Vigenère.

Writing down the possible decodes we see that the first line must read OWING and this makes the other lines read CONDI, ITHAS, EIMPO, ETOIM, ALCUL, MACHI, HISIS, EGRET. By filling in the word CONDITIONS the whole can now be decoded.<sup>5</sup>

<sup>5</sup> Solution: Keylength - 10, Key - POIUMOLQNY, Cleartext - OWINGTOWAR CONDITIONS ITHASBECOM EIMPOSSIBL ETOIMPORTC ALCULATING MACHINESXT HISISVERYR EGRETTTABLE



A more accurate argument would run as follows. For the first column, instead of setting up as rival theories the two possibilities that B is the key and that B is not we can set up 26 rival theories that the key is A or B or ... Z, and we may apply the factor principle in the form:-

$$\begin{aligned} & \frac{\text{A posteriori probability of key A}}{\text{A priori probability of key A} \times \text{Probability of getting the given column with key A}} \\ &= \frac{\text{A posteriori probability of key B}}{\text{A priori probability of key B} \times \text{Probability of getting the given column with key B}} \\ &= \text{etc.} \end{aligned}$$

The argument to justify this form of factor principle is really the same as for the original form. Let  $q_\beta$  be the *a priori* probability of key  $\beta$ . Then out of  $N$  cases we have  $Nq_\beta$  cases of key  $\beta$ . Let  $\mathcal{P}(\beta, C)$  be the probability of getting the column  $C$  with key  $\beta$ , then we have rejected the cases where we get columns other than  $C$  we find that there are  $Nq_\beta\mathcal{P}(\beta, C)$  cases of key  $\beta$  *i.e.* the *a posteriori* probability of key  $\beta$  is  $KNq_\beta\mathcal{P}(\beta, C)$ , where  $K$  is independent of  $\beta$ .

We have therefore to calculate the probability of getting the column  $C$  with key  $\beta$  and this is simply  $\prod_i \mathcal{P}_{(\alpha_i - \beta + 1)}$ , *i.e.* the product of the frequencies of the decode letters which we get if the key is  $\beta$ .

Since the *a priori* probabilities of the keys are all equal we may say that the *a posteriori* probabilities are in the ratio  $\prod_i \mathcal{P}_{\alpha_i - \beta + 1}$  *i.e.* in the ratio  $\prod_i 26\mathcal{P}_{\alpha_i - \beta + 1}$  which is more convenient for calculation. The final value for the probability is then

$$\frac{\prod_i 26\mathcal{P}_{\alpha_i - \beta + 1}}{\sum_\beta \prod_i 26\mathcal{P}_{\alpha_i - \beta + 1}}.$$

The calculation of the product  $\prod_i 26\mathcal{P}_{\alpha_i - \beta + 1}$  may be done by the method recommended before for

$$\prod_i \frac{25\mathcal{P}_{\alpha_i - \beta + 1}}{1 - \mathcal{P}_{\alpha_i - \beta + 1}}.$$

(The table in Fig 3 was in fact made up for  $\prod_i 26\mathcal{P}_{\alpha_i - \beta + 1}$ . The differences between the two tables would of course be rather slight). The new result is more accurate than the old because of the independence assumption in the original result.

If we only want to know the ratios of the probabilities of the various keys there is no need to calculate the denominator  $\sum_\beta \prod_i 26\mathcal{P}_{\alpha_i - \beta + 1}$ . This denominator has however another importance: it gives us some evidence about other assumptions, such as that the cipher is Vigenère, and that the period is 10. This aspect will be dealt with later (p. )<sup>6</sup>.

## 2.2. A letter subtractor problem

A substitution with the period  $91 \times 95 \times 99$  is obtained by superimposing three substitutions of periods 91, 95, and 99, each substitution being a Vigenère composed of slides of 0, 1, 2, 3, 4, 5, 6, 7, 8, or 9.<sup>7</sup> The three substitutions are known in detail, but we do not know for any given message at what point in the complete substitution to begin. For many messages however we can provide a more or less probable crib. How can we test the probability of a crib before attempting to

<sup>6</sup> Forward reference left unresolved in the manuscript.

<sup>7</sup> Equivalent to keys A to J.

solve it? It may be assumed that approximately equal numbers of slides 0, 1, . . . , 9 occur in each substitution.

The principle of the calculation is that owing to the way in which the substitution is built up, not all slides are equally frequent, *e.g.* a slide of 25 can only be the sum of slides of 9, 8, and 8, or 9, 9, and 7 whilst a slide of 15 can be any of the following

$$\begin{array}{cccc} 9,6,0 & 8,7,0 & 7,7,1 & 6,6,3 \\ 9,5,1 & 8,6,1 & 7,6,2 & 6,5,4 \\ 9,4,2 & 8,5,2 & 7,5,3 & \\ 9,3,3 & 8,4,3 & 7,4,4 & \end{array}$$

A crib will therefore, other things being equal, be more likely if it requires a slide of 15 than if it requires a slide of 25. The problem is to make the best use of this principle, by determining the probability of the crib with reasonable accuracy, but without spending long over it.

We have to find the probability of getting a given slide. To do this we can apply several methods.

- (a) We can produce a long stretch of key by addition and take a count of the resulting slides. This is obviously a very general method, and requires no special mathematical technique. It may be rather laborious, but by interpreting a small count with common sense one can probably get quite good results.
- (b) There are 1000 possible combinations of slides all equally likely *viz.* 000, 001, . . . , 999. We can add up the digits in these and take the remainder on division by 26, and then count the number of combinations giving each of the possible remainders.
- (c) We can make use of a trick which might appear to be rather special, but is really applicable to a multitude of problems. Consider the expression

$$f(x) = (1 + x + x^2 + \dots + x^9)^3.$$

For each possible way of expressing a number  $n$  as the sum of three numbers 0, . . . , 9, say  $n = m_1 + m_2 + m_3$ , there is a term  $x^{m_1}x^{m_2}x^{m_3}$  in  $f(x)$ ,  $x^{m_1}$  coming out of the first factor,  $x^{m_2}$  out of the second, and  $x^{m_3}$  out of the third. Hence the number of ways of expressing  $n$  in the form  $n = m_1 + m_2 + m_3$ , is the coefficient of  $x^n$  in  $f(x)$  *i.e.* in

$$\frac{(1 - x^{10})^3}{(1 - x)^3},$$

or in

$$(1 - 3x^{10} + 3x^{20} - x^{30})(1 - x)^{-3}.$$

Expanding  $(1 - x)^{-3}$  by the binomial theorem

$$\begin{aligned} (1 - x)^{-3} &= 1 + 3x + 6x^2 + 10x^3 + 15x^4 + 21x^5 + 28x^6 + 36x^7 \\ &\quad + 45x^8 + 55x^9 + 66x^{10} + 78x^{11} + 91x^{12} + 105x^{13} \\ &\quad + 120x^{14} + 136x^{15} + 153x^{16} + 171x^{17} + 190x^{18} \\ &\quad + 210x^{19} + 231x^{20} + 253x^{21} + 276x^{22} + 300x^{23} \\ &\quad + 325x^{24} + 351x^{25} + 378x^{26} + 406x^{27} + 435x^{28} + \dots \end{aligned}$$

Now multiply by  $1 - 3x^{10} + 3x^{20} - x^{30}$  and we get

$$\begin{aligned}
 f(x) = & 1 + 3x + 6x^2 + 10x^3 + 15x^4 + 21x^5 + 28x^6 + 36x^7 \\
 & + 45x^8 + 55x^9 + 63x^{10} + 69x^{11} + 73x^{12} + 75x^{13} \\
 & + 75x^{14} + 73x^{15} + 69x^{16} + 63x^{17} + 55x^{18} \\
 & + 45x^{19} + 36x^{20} + 28x^{21} + 21x^{22} + 15x^{23} \\
 & + 10x^{24} + 6x^{25} + 3x^{26} + x^{27}
 \end{aligned}$$

This means to say that the chances of getting totals 0, 1, 2, ... are in the ratio 1, 3, 6, 10, ... The chances of getting remainders of 0, 1, 2, ... on division by 26 are in the ratio 4, 4, 6, 10, 15, ... To get true probabilities these must be divided by their total which is conveniently 1000.

- (d) There are two other methods, both connected with the last method but not relying so much on the special features of the problem. They will be discussed later.<sup>8</sup>

Suppose then that the probabilities have been calculated by one method or the other (as in fact we have done under (c)). We can then estimate the values of cribs. Let us suppose that a possible crib for a message beginning MVHWUSXOWBVMMK was AMBASSADOR so that the slides were 12, 9, 6, 22, 2, 0, 23, 11, 14. The slide of 12 gives us some slight evidence in favour of the crib being right for slides of 12 occur with frequency 0.073 with right cribs, whilst with wrong cribs they occur with frequency only 1/26. The factor in favour of the crib is therefore  $26 \times 0.073$  or about 1.9. A similar calculation may be made for each of the slides, but of course the work may be greatly speeded up by having the values of the factors  $26 C_s/1000$  in half decibans tabulated: here  $C_s$  is the coefficient of  $x^s$  in the above polynomial  $f(x)$ . The table is given below (Fig 6)

1	0	-20
2	25	-16
3	24	-12
4	23	-8
5	22	-6
6	21	-3
7	20	-1
8	19	1
9	18	3
10	17	4
11	16	5
12	15	6
13	14	6

FIGURE 6. Scores in half decibans of the various slides.

Evaluating this crib by means of this table we score

$$6 + 3 - 3 - 6 - 16 - 20 - 8 + 5 + 6 (= -33),$$

*i.e.* the crib is worse by a factor of  $10^{-33/20}$  than it was before *e.g.* if the *a priori* odds of the crib were 2:1 against it becomes 98:1 against. This crib was in fact made up at random *i.e.* the letters of the cipher text were chosen at random.

<sup>8</sup> No such discussion appears in the manuscript.

Now let us take one made up correctly, *i.e.* really enciphered by the method in question, but with a random chosen key.

N	Y	X	L	N	X	I	Q	H	H
A	M	B	A	S	S	A	D	O	R
13	12	22	11	21	5	8	13	19	16

*(slides)*

This scores 15 so that if it were originally 2:1 against, it now becomes nearly 3:1 on.

Having decided on a crib the natural way to test it is to have a catalogue of the positions in which a given series of slides is obtained if the 91 period component is omitted. We make 91 different hypotheses as to this third component, draw an inference as to what is the part of the slide arising from the components of periods 95 and 99 combined. This we look up in the catalogue. This process is fairly lengthy, and as the scoring of the crib takes only a minute it is certainly worth doing.

### 2.3. Theory of repeats

Suppose we have a cipher in which there are several very long series of substitutions which can be used for enciphering a message, but that one may sometimes get two messages enciphered with the same series of substitutions (or possibly, the series of substitutions for one message being those for another with some at the beginning omitted). In such a case let us say that the messages *fit*, or that they fit at such and such a distance, the distance being the number of substitutions which have to be omitted from the one series to obtain the other series. One will frequently want to know whether two messages fit or not, and we may find some evidence about this by examining the repeats between them.

By the repeats between them I mean this. One writes out the cipher texts of the two messages with the letters which are thought to have been enciphered with the same substitution under one another. One then writes under these messages a series of letters O and X, an O being written where the cipher texts differ and an X where they agree. The series of letters O and X will begin where the second message begins and end where the first to end ends. This series of letters O and X may be called the repetition figure. It may be completed by adding at the ends an indication of how many letters there are which do not overlap, and which message they belong to.

As an example:

```
GFRLIKQGVBMILAFIXMMOROGBYSKYXDAZCHMUMRKBZLDDLDDOHCMTIPRSD
      VLOVDYQCEJSOPYGBMBKYXDAZNBFIPTFCXDOD
      8X000000000X00XX00XXXXXX000000000XOX11
```

On the whole one expects that a fit is more likely to be right the more letters X there are in the repetition figure, and that long series of letters X are especially desirable. This is because it would not be very unusual for two fairly common words to lie directly under one another when the clear texts are written out, thus

```
THEMAINCONVOYWILLARRIVE ...
ALLCONVOYSMUSTREPORT ...
X00XXXXX00000X0000 ...
```

If the corresponding cipher texts really fit, *i.e.* if the letters in the same column are enciphered with the same substitution, then the condition for an  $X$  in the repetition figure of the cipher texts is that there be an  $X$  in the repetition figure of the corresponding clear text. Now series of several consecutive letters  $X$  can occur quite easily as above by two identical words coming under one another, or by such combinations as

ITISEASIERTOTEACHTHANALGEBRA . . .  
 THERAINWASSUCHTHATHECOULD . . .  
 OOOOOOOOOOOOXXXXXOOOOOOOO . . .

if the messages really fit, but if not they can only occur by complete coincidence. One therefore tends to believe that there is a fit when one gets such series of letters  $X$ . As regards single cases of  $X$  the value of them is not so clear, but one can see that if  $\mathcal{P}_\alpha$  is the frequency of letters  $\alpha$  in plain language then the frequency of letters  $X$  as a whole in comparison of plain language with plain language is  $\sum_\alpha \mathcal{P}_\alpha^2$ , whilst for wrong fits of cipher text it is  $1/26$  which is necessarily less. Given a sufficiently long repetition figure one should therefore be able to tell whether it is a fit or not simply by counting the letters  $X$  and  $O$ .

So much is well known. The real point of this section is to show these ideas can be developed into an accurate method of estimating the probabilities of fits.

**2.3.1. Simple form of theory.** The complete theory takes account of the various possible lengths of repeat. As this theory is somewhat complicated it will be as well to give first two simplified forms of the theory. In both cases the simplification arises by neglecting a part of the evidence. In the first simplified form of theory we neglect all evidence except the number of letters  $X$  and the number of letters  $O$ . In the other simplified form the evidence is the number of series of (say) four consecutive letters  $X$  in a repetition figure.

When our evidence is just the number of times  $X$  occurs in the repetition figure, ( $n$  let us say) and the length of the repetition figure ( $N$  say), then the factor in favour of the fit is

$$\frac{\text{Probability of a right repetition figure of length } N \text{ and } n \text{ occurrences of } X}{\text{Probability of a wrong repetition figure of length } N \text{ having } n \text{ occurrences of } X}$$

As an approximation we may assume that the numerator of this expression has the same value as if the right repetition figures were produced letter by letter by independent random choices, with a certain fixed probability of getting an  $X$  at each stage. This probability will have to be  $\beta = \sum_\alpha \mathcal{P}_\alpha^2$ . The numerator is then

$$\begin{aligned} & (\text{Number of repetition patterns with length } N \text{ and } n \text{ occurrences of } X) \\ & \times (\text{Probability of getting a given such repetition pattern by the process just mentioned}), \end{aligned}$$

which we may write as  $R(N, n)Q(N, n)$ . Now let us denote by  $y_i$  the  $i$ th symbol of the given repetition pattern and put  $\tau_x = \beta$  and  $\tau_0 = 1 - \beta$ . Then  $Q(N, n)$ , the probability of getting the repetition pattern is  $\prod_{i=1}^N \tau_{y_i}$  which simplifies to  $\beta^n (1 - \beta)^{N-n}$ . We may do a similar calculation for the denominator, but here we must take  $\beta = 1/26$  since all letters occur equally frequently in the cipher. The denominator is then

$$R(N, n) \left(\frac{1}{26}\right)^n \left(\frac{25}{26}\right)^{N-n}.$$

In dividing to find the factor for the fit  $R(N, n)$  cancels out, leaving

$$(26\beta)^n \left( \frac{26}{25} (1 - \beta) \right)^{N-n}.$$

In other words we score a factor of  $26\beta$  for an X and a factor of  $(26/25)(1 - \beta)$  for an O. More convenient is to regard it as  $10 \log_{10} [(25\beta)/(1 - \beta)]$  decibans for an X and  $10 \log_{10} [(26/25)/(1 - \beta)]$  per unit length of repetition figure (*per unit overlap*).

An alternative argument, leading to the same result, runs as follows. Having decided to neglect all evidence except the overlap and the number of repeats we pretend that nothing else matters, *i.e.* that the form of the figure is irrelevant. In this case we can regard each letter of the repetition figure as independent evidence about the fit. If we get an X the factor for the fit is

$$\frac{\text{Probability of getting an X if the fit is right}}{\text{Probability of getting an X if the fit is wrong}},$$

*i.e.*  $\beta/(1/26)$ . Similarly the factor for an O is  $(1 - \beta)/(25/26)$ .

In either form of argument it is unnecessary to calculate the number  $R(N, n)$ . In this particular case there is no particular difficulty about about it: it is the binomial coefficient. In some similar problems this cancelling out is a great boon, as we might not be able to find any simple form for the factor which cancels. The cancelling out is a normal feature of this kind of problem, and it seems quite natural that it should happen when we think of the second form of argument in which we think of the evidence as consisting of a number of independent parts.

The device of assuming, as we have done here, that the evidence which is not available is irrelevant can often be used and usually leads to good results. It is of course not supposed that the evidence really is irrelevant, but only that the error resulting from the assumption when used in this kind of way is likely to be small.

**2.3.2. Second simplified form of theory.** In the second simplified form of theory we take as our evidence that a particular part of the repetition figure is OXXXXO (say, or alternatively OXXXXO say). The factor is then

$$\frac{\text{Frequency of OXXXXO in right repetition figures}}{\text{Frequency of OXXXXO in wrong repetition figures}}.$$

The denominator is

$$\left( \frac{1}{26} \right)^4 \left( \frac{25}{26} \right)^2,$$

and the numerator may be estimated by taking a sample of language hexagrams and counting the number of pairs that have the repetition figure OXXXXO. The expectation of the number of such pairs is the sum for all pairs of the probabilities of those pairs having the desired repetition figure *i.e.* is the number of such pairs (*viz*  $N(N - 1)/2$  where  $N$  is the size of the sample) multiplied by the frequency of OXXXXO repetition figures. This frequency may therefore be obtained by division if we equate the expected number of these repetition figures to the actual number.

**2.3.3. General form of theory.** It is not of course possible to have statistics of every conceivable repetition figure. We must make some assumptions to reduce the variety that need to be considered. The following assumption is theoretically very convenient, and also appears to be a very good approximation.

The probability of repeats at two points known to be separated by a point where there is known to be no repeat are independent.

We may also assume that the probability of a repeat is independent of anything but the repetition figure in this neighbourhood. (We may however as a refinement produce different positions in a message). We can therefore think of repetition figures as being produced by selecting the symbols of the figure consecutively, the probability of getting an X at each stage being determined by the repetition figure from the point in question back as far as the last 0. Sometimes this will take us back as far as the beginning of the message, and will include the number telling us how many more letters there are which do not repeat at all. We need in practice only distinguish two cases, where this number is 0 and when it is more. We may also neglect the question as to which message occurs first. We therefore have to distinguish the following cases

0	$a_0$	some	$b_0$	none	$c_0$
OX	$a_1$	some X	$b_1$	none X	$c_1$
OXX	$a_2$	some XX	$b_2$	none XX	$c_2$
OXXX	$a_3$	some XXX	$b_3$	none XXX	$c_3$
...		...		...	

The entries  $a_0, a_1, b_0, etc.$  opposite the repetition figures are the notations we are adopting for the probability of getting another X following such a figure. Strictly speaking we should also bring in a notation for the probability of the message coming to an end after any given repetition figure. As the repeats at the end of a comparison do not appear to behave very differently from those in the main part of the message I shall neglect this complication by assuming that the probability of getting an 0 added to the probability of getting an X is 1, and that afterwards one cuts off the end of the series arbitrarily.

Let us calculate the factor for the repeat figure<sup>9</sup>

none X X X X 0 | 0 | 0 | X 0 | X X X 0 | 0 | X X | some

none	X	X	X	X	0
$c_0$	$c_1$	$c_2$	$c_3$	$1 - c_4$	
1/26	1/26	1/26	1/26	25/26	

0	0	X	0
$1 - a_0$	$1 - a_0$	$a_0$	$1 - a_1$
25/26	25/26	1/26	25/26

X	X	X	0
$a_0$	$a_1$	$a_2$	$1 - a_3$
1/26	1/26	1/26	25/26

0	X	X	some
$1 - a_0$	$a_0$	$a_1$	
25/26	1/26	1/26	

<sup>9</sup> In the manuscript, Turing squeezes the following block into three lines by spilling into the document margins and use of pen and ink. The typeset equivalent is unreadable because of the tiny fonts needed to represent the text. The repeat figure is as in the manuscript, the two line following it are broken into four separate tables in order to present the original material in a legible fashion. The vertical lines of the manuscript are the single vertical lines in the repeat figure, and the double vertical line of the tables.

Underneath each symbol has been written the probability that one would get that symbol, knowing the ones which precede, both for the case of a right and of a wrong repetition figure. The factor for the fit is the product of the first row divided by the product of the second. It is convenient to split this up as indicated by the vertical lines into the product of

$$\frac{c_0 c_1 c_2 c_3 (1 - c_4)}{(1/26)^4 \times (25/26)},$$

$$\frac{1 - a_0}{(25/26)}, \quad \text{- occurring three times,}$$

$$\frac{a_0(1 - a_1)}{(1/26) \times (25/26)},$$

$$\frac{a_0 a_1 a_2 (1 - a_3)}{(1/26)^3 \times (25/26)},$$

$$\frac{a_0 a_1}{(1/26)^2},$$

and this product may be put into the form of the product of

$$\frac{c_0 c_1 c_2 c_3 (1 - c_4)}{(1/26)^4 \times (25/26)} \times \left( \frac{1 - a_0}{(25/26)} \right)^{-5},$$

- which we call the factor for an initial tetragramme repeat level,

$$\frac{a_0(1 - a_1)}{(1/26) \times (25/26)} \times \left( \frac{1 - a_0}{(25/26)} \right)^{-2},$$

- the factor for a single repeat,

$$\frac{a_0 a_1 a_2 (1 - a_3)}{(1/26)^3 \times (25/26)} \times \left( \frac{1 - a_0}{(25/26)} \right)^{-4},$$

- the factor for a trigramme,

$$\frac{1 - a_0}{1 - a_2},$$

- the correction for a final bigramme,

$$\left( \frac{1 - a_0}{(25/26)} \right)^{16},$$

- the factor for an overlap of 16,

$$\frac{a_0 a_1 (1 - a_2)}{(1/26)^2 \times (25/26)} \times \left( \frac{1 - a_0}{(25/26)} \right)^{-3},$$

- the factor for a trigramme.

We shall neglect the correction for a final bigramme (or whatever it may be). It is in any case rather small, and vanishes if the repetition figure ends with 0; also with our conventions the whole question of the ends of repetition figures has been left rather in doubt.



Now let us put<sup>10</sup>

$$\begin{aligned} a_0 a_1 \dots a_r (1 - a_{r+1}) &= k_r, \\ b_0 b_1 \dots b_r (1 - b_{r+1}) &= j_r, \\ c_0 c_1 \dots c_r (1 - c_{r+1}) &= i_r. \end{aligned}$$

The values of the  $i_r$  can be obtained as follows. We take a number of plain language messages and leave out two or three words at the beginning. Then combine the messages to form one long message; this message may be made to *eat its own tail* i.e. it may be written round a circle. If the message were compared with itself in every possible position, except level, we should expect to get repetition figures which when divided up as shown by vertical lines after each  $\emptyset$ , containing  $(N(N-1)/2)k_r$  ( $= N_r$ ) parts which consist of  $r$  symbols  $\emptyset$ , or as we may say  $N_r$  actual  $r$ -gramme repeats, where  $h$  is the probability of an  $\emptyset$ .

The values of  $N_r$  can be calculated given the *apparent number of  $r$ -gramme repeats*  $M_r$  for each  $r$ . This apparent number of  $r$ -gramme repeats is the number of series of  $r$  consecutive symbols  $X$  in the repetition figures regardless of what precedes or follows the series.

By considering the ways in which an actual repeat can give rise to the apparent repeat of various lengths we see that

$$M_r = N_r + 2N_{r+1} + 3N_{r+2} + \dots,$$

and therefore

$$M_r - M_{r+1} = N_r + N_{r+1} + N_{r+2} + \dots,$$

and

$$(M_r - M_{r+1}) - (M_{r+1} - M_{r+2}) = N_r.$$

The calculation of  $j_r$  may perhaps best be done by comparing the beginners of a number of messages with the long circular message, and the values of  $i_r$  by comparing the beginners among themselves. A similar technique of actual and apparent numbers of repeats can be used. I shall not go into this in detail. The formulae required may now be assembled.

$\mu_r$  = decibanage for an  $r$ -gramme repeat,

$\gamma$  = negative decibanage for unit overlap,

$S_{\beta,r}$  = number of occurrences in the statistics of the  $r$ -gramme  $\beta$ ,

$N$  = total number of letters in the statistics.

Then if

$$M_r = \sum_{\beta} \frac{S_{\beta,r} (S_{\beta,r} - 1)}{2},$$

$$N_r = M_r - 2M_{r+1} + M_{r+2},$$

$$L = \frac{N(N-1)}{2},$$

$$k_r = \frac{N_r}{Lh}.$$

---

<sup>10</sup> The manuscript has a pencilled note beside  $k_r$  indicating it is to be read as  $k_{r+1}$ . We presume that this also means that  $j_r$  should be  $j_{r+1}$ , and  $i_r$  should be  $i_{r+1}$ . However, these are not indicated and no changes are made in the subsequent text. We leave the text unchanged

$h$  may be calculated as follows. From the identity

$$(1 - a_0) + a_0(1 - a_1) + a_0a_1(1 - a_2) + \cdots = 1,$$

we get

$$\begin{aligned} k_0 + k_1 + k_2 + \cdots &= 1, \\ \therefore \frac{L - M_1}{Lh} &= 1, \\ (1 - a_0) = k_0 &= \frac{N_0}{L - M_1} = \frac{L - 2M_1 + M_2}{L - M_1}, \\ \mu_r &= 10 \log_{10} \left( \frac{26^{r+1} k_r}{25} \right) + (r + 1) \nu, \\ \nu &= -10 \log_{10} \left( \frac{26(1 - a_0)}{25} \right). \end{aligned}$$

## 2.4. Transposition ciphers

**2.4.1. A probability problem.** In making calculations about substitution ciphers we have often found it useful to treat the plain language as if it were produced by independent choices for the letters, using certain fixed frequencies with which the letters are chosen. Our method for Vigenère and one of the simplified forms of repeat theory could be based on this sort of assumption. With a transposition cipher however such an assumption would be useless or worse than useless, for it would result in the conclusion that all transpositions were equally likely. We have therefore to take a slightly less crude assumption, and the one which suggests itself is that the letters forming the plain language are chosen consecutively, the probability of getting a particular letter depending only on what the letter is and what the preceding letter was. It is easily verified that if  $\mathcal{P}_{\alpha\beta}$  is the proportion of bigrammes  $\alpha\beta$  in plain language and  $\mathcal{P}_\alpha$  the frequency of the letter  $\alpha$  then the probability  $q_{\alpha\beta}$  of a letter  $\beta$  following an  $\alpha$  is  $\mathcal{P}_{\alpha\beta}/\mathcal{P}_\alpha$ . The probability of a piece of plain language of length  $L$  letters saying  $\alpha_1\alpha_2 \dots \alpha_L$  is then

$$\mathcal{P}_{\alpha_1} \times q_{\alpha_1\alpha_2} \times q_{\alpha_2\alpha_3} \times q_{\alpha_3\alpha_4} \times \cdots \times q_{\alpha_{L-1}\alpha_L},$$

which may also be written as

$$\mathcal{J}(\alpha_1, \dots, \alpha_L).$$

We may also calculate the probability of a given piece of plain language having certain given letters in given places, the remainder of the message being unspecified. The probability is given by

$$\sum (\xi_1, \dots, \xi_L \text{ consistent with data}) \mathcal{J}(\xi_1, \dots, \xi_L),$$

and if the data is that the known letters are

$$\begin{array}{ccccccc} \cdots & \beta_1 & \cdots & \beta_2 & \cdots & \cdots & \beta_{r-1} & \cdots & \beta_r & \cdots, \\ n_1 \text{ dots} & & n_2 \text{ dots} & & & & & n_r \text{ dots} & & \end{array} \quad (1)$$

it is approximately<sup>11</sup>

$$\prod_r \mathcal{P}_{\beta_r} \cdot \prod_{n_{r+1}=0} \frac{\mathcal{P}_{\beta_r\beta_{r+1}}}{\mathcal{P}_{\beta_r}\mathcal{P}_{\beta_{r+1}}}. \quad (2)$$

<sup>11</sup> The manuscript has as the first term  $\prod_r \beta_r$ , a pencilled annotation indicates that the  $\beta_r$  is to be read as  $\mathcal{P}_{\beta_r}$ . This substitution has been made in the text.

A more or less rigorous deduction of this approximation from the assumptions above is given at the end of the section. For the present let us see how it can be applied. If we have two theories about the transposition of which the one requires the above pattern of letters, and the other brings the same letters in to positions in which no two of them are consecutive, then the factor in favour of the first as compared with the second is

$$\prod_{n_{r+1}=0} \frac{\mathcal{P}_{\beta_r \beta_{r+1}}}{\mathcal{P}_{\beta_r} \mathcal{P}_{\beta_{r+1}}}.$$

We can apply this straightforwardly to the case of a simple transposition by columns. The following text is known to be a simple transposition of a certain type of German text with a key length of not more than 15.<sup>12</sup>

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

To solve this transposition, we may try comparing the first six letters of S A T P T W which we know form part of one column with each other series of six letters in the message, for we know that one such comparison will give entirely bigrammes occurring in the decode. We may try first

S F  
 A A  
 T S  
 P T  
 T A  
 W U

The factor for a transposition which brings these letters together, as compared with one which leaves them apart is

$$\frac{\mathcal{P}_{SF}}{\mathcal{P}_S \mathcal{P}_F} \times \frac{\mathcal{P}_{AA}}{\mathcal{P}_A \mathcal{P}_A} \times \cdots \times \frac{\mathcal{P}_{WU}}{\mathcal{P}_W \mathcal{P}_U}.$$

By using a table of values of

$$20 \log_{10} \left( \frac{\mathcal{P}_{\alpha\beta}}{\mathcal{P}_\alpha \mathcal{P}_\beta} \right),$$

made up for the type of traffic in question, and given to the nearest integer (table of values of  $\mathcal{P}_{\alpha\beta}/(\mathcal{P}_\alpha \mathcal{P}_\beta)$  expressed in half-decibans) we get the product by addition. Such a table is shown in Fig 6. The scores for this particular columns are SF -7, AA -7, TS -2, PT -10, TA -3, WU -13, totalling -36. If we consider this combination as *a priori* about 100:1 against (there are 95 letters in the message) it is *a posteriori* about 3000:1 against.

<sup>12</sup> As for the Vigenère problem above, Turing's statement of the ciphertext is slightly different from that which he scores for decryption. The second line in the ciphertext below begins NLTS, however, this changes to NITS in the scoring example in Figure 7 below. See also the notes accompanying the cleartext.

A	-7	5	10	-4	-5	-7	3	-4	-25	-7	-8	-1	1	3	-20	-6	-2	-1	-6	-6	8	0	-11	-20	-7	-13	A
B	-7	6	-10	-9	9	-13	-6	-13	1	-11	-12	-7	-2	-18	4	-13	-2	1	-13	-15	-1	-16	-12	-18	-4	-12	B
C	-14	-3	-3	-18	-19	-20	-19	27	-21	-10	3	-12	-12	-21	-15	-13	-2	-14	-14	-21	-21	5	-17	-17	-14	-17	C
D	5	-8	-18	-18	4	-6	-13	-16	4	2	-9	-11	-6	-13	4	4	-2	9	-6	-11	2	-2	-9	-10	-5	-4	D
E	-15	4	0	-8	-15	-5	-2	-5	10	-8	-14	1	-1	5	-22	-6	-3	8	-2	-6	-5	-6	-11	-13	-8	-4	E
F	-2	-11	-20	-9	-2	10	-3	-15	-12	4	-2	1	-8	-11	6	0	-3	-8	-8	-1	9	-8	-8	1	-2	-1	F
G	-1	-8	-18	-5	8	-10	-2	-13	-10	2	6	-3	-13	-11	-10	-14	-3	0	-3	-9	-2	-7	-7	1	2	-2	G
H	1	-10	-12	-8	4	-5	-11	-12	-2	5	-10	2	-3	-10	-2	-2	-2	4	-1	9	-11	-7	-4	-7	-15	-8	H
I	-14	-4	10	-10	0	-1	2	-18	-17	-6	-5	0	-1	9	-17	-7	-1	-10	-1	4	-19	-7	-16	-3	-9	-4	I
J	3	3	-4	1	-3	0	-1	2	-6	14	1	-3	1	-7	-3	7	-2	-12	1	-8	-4	5	-2	0	9	-12	J
K	-2	-9	-12	3	-3	1	-7	-9	-9	4	20	-2	1	-14	-15	-13	7	-6	-5	0	0	-3	-6	-17	-5	-8	K
L	6	0	-6	2	-4	-7	-1	-15	1	-3	-14	8	-4	-2	-2	-5	8	-18	-5	-5	2	-1	-10	3	-5	-3	L
M	6	-1	-17	-6	1	-9	-6	-5	5	1	-10	-14	15	-14	0	-2	-2	-14	-6	-5	-11	-4	-7	3	-6	2	M
N	-1	-8	-18	10	-6	3	11	-9	-8	2	-6	-11	-5	-7	-2	-9	2	-20	6	-6	4	-3	-6	3	0	-5	N
O	-10	-8	-10	-6	-3	4	-6	-13	-18	0	-1	2	6	0	1	2	-2	9	0	2	-18	3	-11	6	-6	-2	O
P	2	-7	-13	-13	4	-3	-14	-5	-8	3	-12	0	-2	-8	6	8	-2	5	-15	-10	5	-12	-12	-13	-11	-1	P
Q	-3	-2	-2	-2	-3	-3	-3	-2	-2	-1	-3	-2	-2	-3	-2	-2	3	-3	-3	-3	13	-2	-2	-2	-2	-2	Q
R	2	3	-3	5	2	0	-6	-10	-2	2	-1	-6	-2	-9	-6	1	-1	-11	-1	2	-1	-2	0	1	-1	2	R
S	-3	-1	11	-2	-4	-7	-13	-12	2	-6	1	-9	-10	-7	-5	8	-3	-19	5	7	-8	5	-9	-15	1	4	S
T	3	-3	-14	-7	5	-3	-11	-11	-4	1	-1	-4	-5	-9	-2	-7	-3	-2	-2	5	-5	-3	-4	2	-3	9	T
U	-11	-4	-3	-15	0	5	-6	-2	-26	-12	-16	10	-2	8	-11	3	2	-3	-3	-11	-2	-9	-11	-12	-9	-20	U
V	-13	-16	-15	-18	2	-15	-16	-5	12	-10	-7	-16	-15	-17	14	-12	-2	-17	-18	-8	-11	10	-14	-4	8	-8	V
W	-1	-17	-17	-17	4	-18	-18	-18	2	-10	-9	-13	-10	-20	21	-12	-2	-19	-20	-19	-13	-14	-3	-16	-6	-16	W
X	1	1	-13	6	1	0	-4	-13	-14	3	0	-12	2	-10	-12	-2	8	-15	-6	-4	-5	9	1	10	-6	3	X
Y	-11	-1	-14	7	-5	-4	-9	-9	-11	-9	0	-9	-6	-5	-4	-11	-2	-11	-6	-6	-7	9	-2	-13	25	-2	Y
Z	-13	-13	-17	-13	-4	-12	-8	-14	-7	-3	-10	-11	-17	-16	-1	-5	-2	-20	-13	-10	8	-8	27	0	-6	-2	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z		

FIGURE 6. Exclusive bigramme scores in half decibans, *i.e.*  $20 \log_{10} \left( \frac{P_{\alpha\beta}}{P_{\alpha}P_{\beta}} \right)$ , for a certain kind of German traffic.

-36	S			X	F	-22		X	E	-36
X	A			13	C	37		X	O	-55
X	T			X	U	-23		X	T	-8
X	P			X	I	-1		X	U	-40
X	T			X	E	-2		X	S	-18
X	W		X	X	B	-49		X	H	-44
X	S	-10	X	X	O	-44		X	L	-43
X	F	-36	X	X	E	-59		X	E	-38
-19	A	-26	X	-22	Y	-31		X	S	-32
X	S	16	X	X	Q	-25		<b>7</b>	B	-32
X	T	-6	6	X	H	8		X	F	-35
-10	A	18	-7	X	G	-19		-17	B	-40
X	U	0	X	-26	T	-4		-38	R	-67
X	T	9	7	X	J	<b>-4</b>		X	N	-42
X	E	-22	-8	X	T	4			D	-58
X	E	-24	X	X	E	6		X		-40
X	A	-39	X	X	E	-32		G		-20
X	I	-25	-10	X	F	-38		N		-37
X	E	-27	-15	X	I	1		J		-25
X	U	-23	X	X	E	-52		H		-42
X	F	-43	X	-11	O	-14		U		
X	H	-27	X	X	R	-46		A		
X	W	-53	X	-13	T	-19		N		
X	T	-38	-22	X	A	-26		W		
X	J	-60	X	X	R	-45		R		
X	T	-52	-25	X	U	-47				
X	D	-37	X	X	R	-87				
X	D	-44	-7	X	N	-54				
X	G	-45	X	X	L	-33				
7	C	2	X	X	N	-16				
X	N	-56	X	-13	N	-11				
X	I	-22	X	-5	N	17				
X	T	-11	-15	-24	N	-36				
X	S	-19	X	X	A	-25				
X	E	-18	<b>27</b>	X	I	-40				
	F				E					
	C				O					
	U				T					
	I				U					
	E				S					

FIGURE 7. Scoring the matching of columns in a simple transposition. Correct matchings noted.

Similar scoring may be done for every possible comparison of S A T P T W with six consecutive letters of the message. The comparison may be made both with S A T P T W as earlier and as later column; one may also use the last six letters of the message H U A N W R.

The results of doing this are shown in Fig 7. The message has been written out vertically. The first columns of figures after the message gives the score for S A T P T W as earlier column, entered against the first letter of the later column, *e.g.* the -36 as calculated above gets entered against the F of F A S T A U. The second column after the message consists of the scores for H U A N W R as first column [and the column before the message gives scores for H U A N W R as second column].<sup>13</sup> One of these columns has been worked out in detail but in the other two crosses have been put in where the scores are very bad.

The scores which eventually turned to to be right are ringed. The fourth comparison, which did not have to be done scored very badly *viz.* -27. Amongst the good scores which were wrong there was one score of 37. It was not difficult to see that this one was wrong as most of the score came from W O with requires Z to precede it, and there was no Z in the message. Apart from this fact the comparison was about evens, although if we take into account the fact that there was no better score it would be better.<sup>14</sup> [We have already had a case of this kind of thing in connection with Vigenère; if the various positions are *a priori* equally likely and the factors are  $f_1, f_2, \dots, f_N$  then the value  $f_r / \sum f_i$  for the probability of the  $r$ th alternative is better than  $(f_r/N) / (1 + f_r/N)$ ].

**2.4.2. The Probability Formula.** (Semi) Rigorous deduction of the formula (2) on page 20. (This is something of a digression).

The probability of a piece of plain language coinciding where necessary with the data (1) on page 20 is

$$\mathcal{P}_{\beta_1} \mathcal{T}_{n_2, \beta_1 \beta_2} \mathcal{T}_{n_3, \beta_2 \beta_3} \cdots \mathcal{T}_{n_m, \beta_{m-1} \beta_m},$$

where

$$\mathcal{T}_{n, \alpha\beta} \text{ is } \sum_{\eta_1 \eta_2 \cdots \eta_n} q_{\alpha \eta_1} q_{\eta_1 \eta_2} \cdots q_{\eta_n \beta},$$

since

$$\sum_{\eta_1 \cdots \eta_{n_1}} \mathcal{P}_{\eta_1} q_{\eta_1 \eta_2} \cdots q_{\eta_{n_1} \beta_1} = \mathcal{P}_{\beta_1}.$$

We can put

$$\mathcal{T}_{n, \alpha\beta} = (\mathcal{Q}^{n+1})_{\alpha\beta},$$

where  $\mathcal{Q}$  is the matrix whose  $\alpha\beta$  coefficient is  $q_{\alpha\beta}$ . The formula (2) on page 20 would then be accurate if we could say that for  $n > 0$ ,

$$(\mathcal{Q}^{n+1})_{\alpha\beta} = \mathcal{P}_{\beta}.$$

This is not true, but it is true that except for very special values for  $q_{\alpha\beta}$ ,

$$(\mathcal{Q}^n)_{\alpha\beta} \rightarrow \mathcal{P}_{\beta}, \text{ as } n \rightarrow \infty,$$

<sup>13</sup> ... and the column to end of sentence, has the note in pencil: *I doubt it - S.W.*

<sup>14</sup> Using Turing's scoring recommendations and a key length of 12 with sequence 5, 11, 8, 7, 3, 10, 6, 12, 9, 4, 1, 2, a cleartext emerges: BNT0 SJJ ALBA RFJ STATT IN OST B HEUTE DEN ETA RUF5 PEDUNYAR NACHT FGFNQUDNUL WICH AHTR X WIESEN WI GEN GRESFOITE TE. With Turing's original statement of the ciphertext, as noted above, GRESFOITE becomes GRESFOLTE. Turing scores for the I and not for L, although it makes no differences in the decision to align bigrams.

and the convergence is rather rapid.

To prove this I shall assume that the eigenvalues of  $\mathcal{Q}$  are all different in modulus. In this case we can find a matrix  $\mathcal{U}$  with unit determinant, such that  $\mathcal{U}^{-1}\mathcal{Q}\mathcal{U}$  is in the diagonal form

$$\mathcal{M} = \mathcal{U}^{-1}\mathcal{Q}\mathcal{U} = \begin{pmatrix} \mu_1 & 0 & 0 & \cdots & 0 \\ 0 & \mu_2 & 0 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & 0 & \mu_{25} & 0 \\ 0 & \cdots & 0 & 0 & \mu_{26} \end{pmatrix},$$

since  $\mathcal{Q}\mathcal{U} = \mathcal{U}\mathcal{M}$  we have

$$\sum_{\gamma} q_{\alpha\gamma} u_{\gamma\beta} = \sum_{\xi} u_{\alpha\xi} m_{\xi\beta},$$

*i.e.*

$$\sum_{\gamma} q_{\alpha\gamma} u_{\gamma\beta} = \mu_{\beta} u_{\alpha\beta}.$$

That is, for each  $\beta$ ,  $u_{\alpha\beta}$  provides a solution of

$$\sum_{\gamma} q_{\alpha\gamma} l_{\gamma} = \mu l_{\alpha}, \quad (3)$$

with  $\mu = \mu_{\beta}$ . Conversely if we have any solution of (3) then  $\mu = \mu_{\vartheta}$ ,  $l_{\alpha} = k u_{\alpha\vartheta}$  for some  $k, \vartheta$  and all  $\alpha$ , for as  $\mathcal{U}$  is non singular we can find numbers  $c_{\gamma}$  such that

$$l_{\alpha} = \sum_{\gamma} u_{\alpha\gamma} c_{\gamma} \text{ for all } \alpha,$$

and then substituting in (3) we get

$$\sum_{\gamma, \delta} q_{\alpha\gamma} u_{\gamma\delta} c_{\delta} = \mu \sum_{\delta} u_{\alpha\delta} c_{\delta},$$

*i.e.*

$$\sum_{\delta} (\mu_{\delta} c_{\delta} - \mu c_{\delta}) u_{\alpha\delta} = 0.$$

Which, since  $\mathcal{U}$  is nonsingular implies  $\mu = \mu_{\delta}$  or  $c_{\delta} = 0$  for all  $\delta$ .

As the series  $\mu_1, \dots, \mu_{26}$  are all different there is only one value  $\vartheta$  of  $\delta$  for which  $\mu = \mu_{\delta}$  and so  $l_{\alpha} = c_{\vartheta} u_{\alpha\vartheta}$  for all  $\alpha$ . Now putting  $l_{\alpha} = 1$  for all  $\alpha$  we see that one member of the series  $\mu_1, \dots, \mu_{26}$  is 1, for (3) is certainly satisfied.

I shall prove that the remaining eigenvalues satisfy  $|\mu| \leq 1$ . We first prove that if  $\mu \neq 1$  then  $\sum p_{\alpha} l_{\alpha} = 0$ . This follows by multiplying (3) on each side by  $\mathcal{P}_{\alpha}$  and summing. Since

$$q_{\alpha\beta} = \frac{\mathcal{P}_{\alpha\beta}}{\mathcal{P}_{\alpha}} \text{ and } \sum_{\alpha} \mathcal{P}_{\alpha\beta} = \mathcal{P}_{\beta},$$

we get

$$\sum_{\alpha\gamma} q_{\alpha\gamma} l_{\gamma} = \sum p_{\gamma} l_{\gamma} = \mu \sum p_{\alpha} l_{\alpha},$$

which implies

$$\mu = 1 \text{ or } \sum p_{\alpha} l_{\alpha} = 0.$$

Next we show that each  $\mu$  for which  $|\mu| > 1$  is real and positive. Let  $l_\alpha$  satisfy (3) with  $|\mu| > 1$ ; then the eigenvalue for  $\bar{l}_\alpha$  is  $\bar{\mu}$  and so

$$\sum_{\beta} (\mathcal{Q}^r)_{\alpha\beta} (1 + \varepsilon (l_\beta + \bar{l}_\beta)) = 1 + 2\varepsilon \Re \mu^r l_\alpha.$$

If  $\varepsilon > 0$  has been chosen so small that  $\Re \varepsilon l_\beta > -1/2$  for all  $\beta$  then the L.H.S. is positive for the coefficients in the matrix are positive, whereas the R.H.S. is negative for suitably chosen  $\mu$ , unless  $l_\alpha = 0$ . If now  $\mu > 1$  we may take it that  $l_\alpha$  is real for each  $\alpha$ . As it must satisfy  $\sum p_\alpha l_\alpha = 0$  it is negative for some  $\alpha$ , but then

$$\sum_{\beta} (\mathcal{Q}^r)_{\alpha\beta} (1 + \varepsilon l_\beta) = 1 + \varepsilon \mu^r l_\alpha,$$

and if  $\varepsilon$  is chosen so that  $1 + \varepsilon l_\beta > 0$  for all  $\beta$  the L.H.S. is positive whereas the R.H.S. is negative for sufficiently large  $r$ .

All the eigenvalues therefore satisfy  $|\mu| \leq 1$  as the eigenvalues are all different in modulus this means that  $|\mu| < 1$  except for one value of  $\mu$ . Then as  $r \rightarrow \infty$ ,  $\mathcal{M}^r$  tends to a matrix which has only one element different from 0, and that a 1 on the diagonal, say in position  $\sigma\sigma$ .

Calling this matrix  $\mathcal{X}_\sigma$  the series of matrices  $\mathcal{Q}^r$  tends to the matrix  $\mathcal{U}^{-1} \mathcal{X}_\sigma \mathcal{U}$ . This matrix is the one and only one  $\mathcal{Y}$  which satisfies  $\mathcal{Y}\mathcal{Q} = \mathcal{Y}$ ,  $\mathcal{Y}^2 = \mathcal{Y}$ ,  $\mathcal{Y} \neq 0$  and is therefore the one whose  $\alpha\beta$  coefficient is  $\mathcal{P}_\beta$ .

**2.4.3. Another probability problem.** There is another probability problem that arises in connection with simple transpositions. With a message of length  $L$ , and a key length of  $k$  what is the probability that the  $m$ th letter will be at the bottom of a column? Let  $D$  be the length of the short columns *i.e.*  $D = [L/K]$ , and let  $E = L - DK$ . Then if the  $m$ th letter is at the bottom of the  $w$ th column we must have

$$\frac{m}{D+1} \leq w \leq \frac{m}{D},$$

and there will be  $(D+1)w - m$  short and  $m - Dw$  long columns among these first  $w$  columns. There are<sup>15</sup>

$$\binom{w}{m - Dw} \binom{K - w}{E - m + Dw}$$

ways in which the short and long columns can be arranged consistently with this, and altogether  $\binom{K}{E}$  ways in which the columns can be arranged, so that the probability of the  $m$  the letter being at the bottom of a column is

$$\sum_{(m/D+1) \leq w \leq (m/D)} \binom{w}{m - Dw} \binom{K - w}{E - m + Dw} \bigg/ \binom{K}{E}.$$

There will normally be very few terms in the sum.

<sup>15</sup> Turing is using Binomial Coefficient notation in this section;

$$\binom{n}{k} = C(n, k) = \frac{P(n, k)}{P(k, k)} = \frac{n!}{(n-k)!k!}$$



Let us take the case of the message of length 133 and consider the 45th letter, assuming the key length is between 10 and 20 (inclusive).  $L_O^B$ .  $L = 133, m = 45$ .<sup>16</sup>

TABLE 1. Message term values

K	D	E	m/(D+1)	m/D	Comment
10	13	3	3+	3+	<i>no terms</i>
11	12	1	3+	3+	<i>no terms</i>
12	11	1	3+	4+	<i>only terms w = 4, m - Dw = 1</i>
13	10	3	4+	4+	<i>no terms</i>
14	9	7	4+	4+	<i>only terms w = 5, m - Dw = 0</i>
15	8	13	5	4+	<i>only terms w = 5, m - Dw = 5</i>
16	8	5	5	4+	<i>only terms w = 5, m - Dw = 5</i>
17	7	14	5+	4+	<i>only terms w = 6, m - Dw = 3</i>
18	7	7	5+	4+	<i>no terms</i>
19	7	0	0	+	<i>prob. = 0</i>
20	6	13	6+	7+	<i>only terms w = 7, m - Dw = 3</i>

for K = 12 the probability is:  $\binom{4}{1} \times \binom{8}{0} / \binom{12}{1} = \frac{4}{12}$

for K = 14 the probability is:  $\binom{5}{0} \times \binom{9}{7} / \binom{14}{7} = \frac{3}{286} = 0.0105$ ,

for K = 15 the probability is:  $\binom{5}{5} \times \binom{10}{8} / \binom{15}{13} = \frac{3}{7} = 0.428$ ,

for K = 16 the probability is:  $\binom{5}{5} \times \binom{11}{0} / \binom{16}{15} = \frac{1}{4368} = 0.000229$ ,

for K = 17 the probability is:  $\binom{6}{3} \times \binom{11}{11} / \binom{17}{14} = \frac{1}{34} = 0.0307$   
(*Editor - 1/34 is 0.0294.*)

for K = 18 the probability is:  $\binom{6}{3} \times \binom{12}{4} / \binom{18}{7} = \frac{4950}{15912} = 0.311$ ,

for K = 20 the probability is:  $\binom{7}{3} \times \binom{13}{4} / \binom{20}{7} = \frac{35 \times 143}{15504} = 0.323$ .

---

∞

---

<sup>16</sup> The following material, to the end of the paper, is handwritten in the manuscript. It is presented here in two parts, a table of key information and separate supporting probability calculations.