In the simplest (monoalphabetic) type of substitution cipher, we take a permutation of the alphabet in which the plaintext is written, and substitute each symbol by its image under the permutation. The key to the cipher is the permutation used; anyone possessing this can easily apply the inverse permutation to recover the plaintext.

If we take a piece of ordinary English text, ignore spaces and punctuation, and convert all letters to capitals, then the alphabet consists of 26 symbols, and so the number of keys is

\[26! = 403291461126605635584000000.\]

This is a sufficiently large number to discourage anyone making an exhaustive test of all possible keys. (It is approximately equal to the age of the Universe in microseconds!) However, the cipher is usually very easy to break, as we will see.

We can represent a permutation by writing down the letters of the alphabet in the usual order, and writing underneath each letter its image under the permutation. To find the inverse, write the bottom row above the top row, and then sort the columns so that the new top row is in its natural order. For example, the inverse of the permutation

\[
\begin{align*}
\end{align*}
\]

is

\[
\begin{align*}
\end{align*}
\]

The identity permutation is the very simple permutation which leaves each symbol where it is: not much use for enciphering!

\[
\begin{align*}
\end{align*}
\]

Finally, the composition \(g \circ h\) of two permutations is obtained by applying first \(g\) and then \(h\) to the alphabet.
**Definition**  A set $G$ of permutations forms a group if

(a) for all $g, h \in G$, $g \circ h \in G$;

(b) the identity permutation $e$ belongs to $G$;

(c) for every $g \in G$, the inverse permutation $g'$ belongs to $G$.

The order of the group $G$ is the number of permutations it contains.

For example, the set of all permutations of an $n$-element set is a group, called the *symmetric group* of degree $n$ and denoted by $S_n$. Its order is $n!$. The symmetric group $S_n$ is the set of keys for substitution ciphers with an $n$-letter alphabet.

**Caesar cipher**

The simplest possible substitution cipher is the *Caesar cipher*, reportedly used by Julius Caesar during the Gallic Wars. Each letter is shifted a fixed number of places to the right. (Caesar normally used a shift of three places). We regard the alphabet as a cycle, so that the letter following Z is A. Thus, for example, the table below shows a right shift of 5 places.

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

The message “Send a hundred slaves as tribute to Rome” would be enciphered as *Xjsi f mzsiwji xqfajx fx ywngzyj yt Wtrj*. The key is simply the number of places that the letters are shifted, and the cipher is decrypted by applying the shift in the opposite direction (five places back).

Some practical details make the cipher harder to read. In particular, it would be sensible to ignore the distinction between capital and lower case letters, and also to ignore the spaces between words, breaking the text up into blocks of standard size, for example

```
XJSIF MZSIW JIXQF AJXFX YWNGZ YJYT W TRJXX
```

(We have filled up the last block with padding.)

The Caesar cipher is not difficult to break. There are only 26 possible keys, and we can try them all. In this case we would have

```
XJSIF MZSIW JIXQF AJXFX YWNGZ YJYT W TRJXX
YKTJG NATJX KJYRG BKGY Y ZXOHA ZKZUX USKYY
ZLUHK OBUKY LKZSH CLZHZ AYPIB ALAVY VTLZZ
...  
```
Almost certainly only one of the twenty-six lines will make sense, and it is easy to break it into words and discard the padding.

There are other tricks that can be used, which will be important later. As we will see in the next section, in English text, the commonest letter is usually E. Also, the consecutive letters R, S, T, U are common, and are followed by a block V, W, X, Y, Z of relatively uncommon letters. If we can spot these patterns, then we can make a guess at the correct shift. Our example is too short to show much statistical regularity; but (if we assume that the last two Xs are padding) the commonest letter is J, and the letters W, X, Y, Z are common while A, B, C, D, E are rare, so we would guess that the shift is 5 (which happens to be correct). We will look at this again in the next section.

We will in future use the convention that the plaintext is in lower case and the ciphertext in capitals.

A famous modern instance of a Caesar shift was HAL, the rogue computer in the science-fiction story *2001: A Space Odyssey*. The computer’s name is a shift of IBM. (The author, Arthur C. Clarke, denied that he had deliberately done this.)

The Caesar shifts form a group. If the alphabet is $A = \{a_0, a_1, \ldots, a_{q-1}\}$, then the shift by $i$ places can be written as $f_i : a_j \mapsto a_{j+i \mod q}$, and we have

$$f_{i_1} \circ f_{i_2} = f_{i_1+i_2 \mod q},$$

$$f_0 = e,$$

$$f_i' = f_{-i \mod q}.$$

The order of this group is $q$.

**Letter frequencies**

In any human language (and in most artificial languages as well), words are not random combinations of symbols, and so they will show various statistical regularities. For example, in English, the commonest letter is E; in a typical (not too short) piece of English, about 12% of all the letters will be E.

As an example, in the text of *Alice’s Adventures in Wonderland*, by Lewis Carroll (AAIW for short), the frequencies of the letters (ignoring spaces and punctuation) are given in Table 1 (the figure given is the average number of occurrences among 100 letters), in the column labelled “AAIW”. (The figures in the table are the average numbers of occurrences among 100 letters of text.) The columns labelled “Meaker” and “Garrett” are from the books *Cryptanalysis* by Helen Fouché Gaines, and *Making,
Breaking Codes by Paul Garrett. Gaines (whose book was published in 1939) took the numbers from a table by O. P. Meaker; Garrett, on the other hand, simply analysed a megabyte of old email. The French and Spanish statistics are also quoted by Gaines, from tables by M. E. Ohaver, Cryptogram Solving. The last column will be explained later.

Note that even for English text the figures vary, though not too much: in AAIW the most frequent letters, in order, are E, T, A, O, I, H, N, S, R, D, L, U; in Gaines’ table, the order is E, T, A, O, N, I, S, R, H, L, D, U. However, in other languages the order is quite different. For example, in German, the order is typically E, N, I, R, S, A, D, T, U, G, H, O.

Figure 1 shows a histogram of the expected frequencies, together with the actual letter frequencies in the message encrypted by Caesar’s cipher. It is clear by eye that the best fit is obtained if the actual message is shifted five places left.

![Histogram](image)

Figure 1: Expected and actual letter frequencies in Caesar cipher

Pairs of letters (referred to as digrams) also have their characteristic frequencies. Some of the most common in English are given in Table 2. Meaker’s tables, and those of Pratt and Fraprie, are taken from Gaines.

One can also analyse trigrams, or longer sequences. Among the most common trigrams in English are THE, ING, THA, AND, ION.
<table>
<thead>
<tr>
<th>Letter</th>
<th>AAIW</th>
<th>Meaker</th>
<th>Garrett</th>
<th>French</th>
<th>Spanish</th>
<th>Gadsby</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8.15</td>
<td>8.05</td>
<td>7.73</td>
<td>9.42</td>
<td>12.69</td>
<td>10.96</td>
</tr>
<tr>
<td>B</td>
<td>1.37</td>
<td>1.62</td>
<td>1.58</td>
<td>1.02</td>
<td>1.41</td>
<td>2.14</td>
</tr>
<tr>
<td>C</td>
<td>2.21</td>
<td>3.20</td>
<td>3.06</td>
<td>2.64</td>
<td>3.93</td>
<td>2.66</td>
</tr>
<tr>
<td>D</td>
<td>4.58</td>
<td>3.65</td>
<td>3.24</td>
<td>3.38</td>
<td>5.58</td>
<td>4.12</td>
</tr>
<tr>
<td>E</td>
<td>12.61</td>
<td>12.31</td>
<td>11.67</td>
<td>15.87</td>
<td>13.15</td>
<td>0.00</td>
</tr>
<tr>
<td>F</td>
<td>1.86</td>
<td>2.28</td>
<td>2.14</td>
<td>0.95</td>
<td>0.46</td>
<td>2.15</td>
</tr>
<tr>
<td>G</td>
<td>2.36</td>
<td>1.61</td>
<td>2.00</td>
<td>1.04</td>
<td>1.12</td>
<td>3.61</td>
</tr>
<tr>
<td>H</td>
<td>6.85</td>
<td>5.14</td>
<td>4.52</td>
<td>0.77</td>
<td>1.24</td>
<td>4.91</td>
</tr>
<tr>
<td>I</td>
<td>6.97</td>
<td>7.18</td>
<td>7.81</td>
<td>8.41</td>
<td>6.25</td>
<td>8.81</td>
</tr>
<tr>
<td>J</td>
<td>0.14</td>
<td>0.10</td>
<td>0.23</td>
<td>0.89</td>
<td>0.56</td>
<td>0.23</td>
</tr>
<tr>
<td>K</td>
<td>1.07</td>
<td>0.52</td>
<td>0.79</td>
<td>0.00</td>
<td>0.00</td>
<td>1.18</td>
</tr>
<tr>
<td>L</td>
<td>4.37</td>
<td>4.03</td>
<td>4.30</td>
<td>5.34</td>
<td>5.94</td>
<td>5.32</td>
</tr>
<tr>
<td>M</td>
<td>1.96</td>
<td>2.25</td>
<td>2.80</td>
<td>3.24</td>
<td>2.65</td>
<td>2.07</td>
</tr>
<tr>
<td>N</td>
<td>6.52</td>
<td>7.19</td>
<td>6.71</td>
<td>7.15</td>
<td>6.95</td>
<td>8.61</td>
</tr>
<tr>
<td>O</td>
<td>7.58</td>
<td>7.94</td>
<td>8.22</td>
<td>5.14</td>
<td>9.49</td>
<td>10.42</td>
</tr>
<tr>
<td>P</td>
<td>1.40</td>
<td>2.29</td>
<td>2.34</td>
<td>2.86</td>
<td>2.43</td>
<td>1.91</td>
</tr>
<tr>
<td>Q</td>
<td>0.19</td>
<td>0.20</td>
<td>0.12</td>
<td>1.06</td>
<td>1.16</td>
<td>0.05</td>
</tr>
<tr>
<td>R</td>
<td>5.02</td>
<td>6.03</td>
<td>5.97</td>
<td>6.46</td>
<td>6.25</td>
<td>4.77</td>
</tr>
<tr>
<td>S</td>
<td>6.05</td>
<td>6.59</td>
<td>6.55</td>
<td>7.90</td>
<td>7.60</td>
<td>6.97</td>
</tr>
<tr>
<td>T</td>
<td>9.93</td>
<td>9.59</td>
<td>9.53</td>
<td>7.26</td>
<td>3.91</td>
<td>8.50</td>
</tr>
<tr>
<td>U</td>
<td>3.22</td>
<td>3.10</td>
<td>3.21</td>
<td>6.24</td>
<td>4.63</td>
<td>4.16</td>
</tr>
<tr>
<td>V</td>
<td>0.78</td>
<td>0.93</td>
<td>1.03</td>
<td>2.15</td>
<td>1.07</td>
<td>0.31</td>
</tr>
<tr>
<td>W</td>
<td>2.49</td>
<td>2.03</td>
<td>1.69</td>
<td>0.00</td>
<td>0.00</td>
<td>2.80</td>
</tr>
<tr>
<td>X</td>
<td>0.13</td>
<td>0.20</td>
<td>0.30</td>
<td>0.30</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>Y</td>
<td>2.11</td>
<td>1.88</td>
<td>2.22</td>
<td>0.24</td>
<td>1.06</td>
<td>3.18</td>
</tr>
<tr>
<td>Z</td>
<td>0.07</td>
<td>0.09</td>
<td>0.09</td>
<td>0.32</td>
<td>0.35</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 1: Letter frequencies
Table 2: Frequencies of common digrams

<table>
<thead>
<tr>
<th>Digraph</th>
<th>AAIW</th>
<th>Meaker</th>
<th>P &amp; F</th>
<th>Garrett</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH</td>
<td>3.23</td>
<td>3.51</td>
<td>3.16</td>
<td>3.18</td>
</tr>
<tr>
<td>HE</td>
<td>3.23</td>
<td>2.51</td>
<td>1.08</td>
<td>2.17</td>
</tr>
<tr>
<td>AN</td>
<td>1.48</td>
<td>1.72</td>
<td>1.08</td>
<td>1.59</td>
</tr>
<tr>
<td>IN</td>
<td>1.89</td>
<td>1.69</td>
<td>1.57</td>
<td>2.59</td>
</tr>
<tr>
<td>ER</td>
<td>1.68</td>
<td>1.54</td>
<td>1.33</td>
<td>1.95</td>
</tr>
<tr>
<td>RE</td>
<td>1.07</td>
<td>1.48</td>
<td>1.25</td>
<td>1.85</td>
</tr>
</tbody>
</table>

As an indication of how these frequencies reflect the language, here are three “random” pieces of text. In each case, in order to split the text into words, a 27-letter alphabet (consisting of the 26 letters and the space character) has been used; any punctuation characters in the original text are regarded as spaces, and a string of spaces is reduced to a single space. In the first piece of text, the computer has generated random text using the same letter and space frequencies as in AAIW. In the second, the digram frequencies have been used; and in the third, trigram frequencies are used. Notice how the random texts resemble the original more closely as longer sequences are used.

Letter frequencies

garynndidbleayir hedryeeabeslt tyt watat vnot sooannaheoynoc hhh ndn e n mom scie cehealiiea yneuries u imn h uootpn eomvet ia ecadehatyba eub e lsrv utl ecnhrmer etwtata nstp thttwttlt ht mh dyatanpbs toihpiteitsttithotrehushilwhlhtaehyto rovtt aget eaeairwru gnat asrl eeri luikghre borelephre hhvdv egnso nodiehia dcoeothgoa tsabns c neo ndnhfbtsont ne cpoed m t old fzl rohuiinirtosthe arm genialendtr hhntn tsmtr osnol ngohne aiaumnie p hhb te t gtt o araswc tak omlhdtaoi er rlumh ceca tlo acnimal tto sosi ah htoe c sty laahsouseshi oae oh aafasth wnsihnaeoawoi aesnhi yb vresptn gas elplteot or anner en s e dfhat tso nmrl te euhdre ltsnsr f reesd s cchtovahns uhtiwalo tahot lrrnt

Digram frequencies

tre wherrltau ar a inor hee ly groove aye abinglothased as an nonnte fin whike it im yon coveng a per weker ligo d ated ay s red ase ous andldrthi i anory acke owhalist the w an thi tuth abinwaly lyton bofforyilenour t n ns art asod h athostugir telidademifure bing hee hedertlirouricell araks edshe capl asove a asino that ar at heldryirry id and aghanorsith anesance
Breaking a substitution cipher

Breaking a cipher is an art; it cannot be done by applying a formula. But there are some rules to follow when doing this job. Here is a partly worked example of breaking a substitution cipher; you should complete the working.

The ciphertext is:

\[
\begin{align*}
RZOLB & \quad QJOWW & \quad QBWIR & \quad DQFQE & \quad VICOB & \quad OKOLR & \quad UVIDW & \quad QFMRO & \quad IVTOH \\
OVZMA & \quad UFURR & \quad UVEWM & \quad DWDBH & \quad UOVYO & \quad RQRZO & \quad UBWRM & \quad TOVRW & \quad RZOSZ \\
ITRQW & \quad COIBQ & \quad DOTUO & \quad VYORQ & \quad ZRORW & \quad MTOVR & \quad BOYRQ & \quad BWIVT & \quad RQRZO \\
WRMTO & \quad VRAIT & \quad OWRIR & \quad MROWC & \quad ZUYZD & \quad QBOHO & \quad BSZIB & \quad TFSML & \quad QVRZO \\
ARZOL & \quad BQJOW & \quad WQBCI & \quad WJUVO & \quad TUIJZO & \quad DOEIV & \quad ZUWRO & \quad IYZUV & \quad EIAUV \\
MROFI & \quad ROQBY & \quad QVRUV & \quad MOTIA & \quad UVMRO & \quad FQVEO & \quad BRZIV & \quad RZOJU & \quad XOTRU \\
AOIVT & \quad WZQMF & \quad TRZUW & \quad ZILLO & \quad VRZOW & \quad RMTOV & \quad RWCZQ & \quad JIUFO & \quad TRQFO \\
IHORZ & \quad OFOYR & \quad MBOBQ & \quad QAUA & \quad OTUIR & \quad OFSCO & \quad BORZO & \quad AWOFH & \quad OWJUV \\
OTUJI & \quad TTURU & \quad QVRZO & \quad LBQJO & \quad WWQBC & \quad IWJUV & \quad OTUJZ & \quad OWZUB & \quad KOTOX \\
LFIUV & \quad UVEIT & \quad UJJUY & \quad MFRLI & \quad WNWEO & \quad QBUJZ & \quad OJIUF & \quad OTRQE & \quad ORRZB \\
QMEZR & \quad ZOWSF & \quad FIDMW & \quad ZOCIW & \quad JUVOI & \quad UJZOF & \quad OJRRZ & \quad OYURS & \quad JQBIT \\
ISCUR & \quad ZQMRR & \quad UOBOY & \quad RQBWJ & \quad OBAUW & \quad WUQVI & \quad VTUJJ & \quad OAJBB & \quad UOTCI \\
WIFFQ & \quad COTQV & \quad FSQVO & \quad TISQJ & \quad JJQBR & \quad ZOLMB & \quad LQWOR & \quad ZOYUR & \quad SJQBU
\end{align*}
\]
We first count the frequencies of the letters. The commonest of the 715 letters, with their frequencies, are given in the table.

<table>
<thead>
<tr>
<th>Letter</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>99</td>
</tr>
<tr>
<td>R</td>
<td>72</td>
</tr>
<tr>
<td>Q</td>
<td>59</td>
</tr>
<tr>
<td>I</td>
<td>50</td>
</tr>
<tr>
<td>U</td>
<td>49</td>
</tr>
<tr>
<td>W</td>
<td>48</td>
</tr>
<tr>
<td>V</td>
<td>45</td>
</tr>
<tr>
<td>B</td>
<td>43</td>
</tr>
<tr>
<td>Z</td>
<td>43</td>
</tr>
</tbody>
</table>

We also notice that RZ is a very common digram, with 23 occurrences. So we might guess the following identifications: O = e, R = t, Z = h. This gives

The other common letters probably include a, i, o and n. Various clues help us to make the right identification. For example, consider the string tQthe, which occurs several times. Here, the is probably either a word or the beginning of a word like then. If this is right, tQ ends a word, and the most likely possibility is that Q = o.

Another clue is that WQ occurs four times in the text. Double letters are not very common in English; ee, ll and ss are the most common, so probably W = s. After a certain amount of guesswork of this sort, we begin to recognise more complicated words, and we find eventually that the substitution is

The professors at Bologna were kept in absolute and even humiliating subservience to their students. They had to swear obedience to the
student rectors and to the student-made statutes, which bore very hardly upon them. The professor was fined if he began his teaching a minute late or continued a minute longer than the fixed time, and should this happen the students who failed to leave the lecture-room immediately were themselves fined. In addition, the professor was fined if he shirked explaining a difficult passage, or if he failed to get through the syllabus; he was fined if he left the city for a day without the rector’s permission, and if he married, was allowed only one day off for the purpose. The city, for its part, took a hand in controlling the professors, and they were forced to take an oath not to leave Bologna in search of more lucrative or less onerous posts.

This description of employment conditions for academics in the Middle Ages is taken from J. D. Knowles, *The Evolution of Mediaeval Thought*.


**Worked example**  Solve the following substitution cipher.

\[
\begin{align*}
\text{Solution:} & \text{ This cipher is surprisingly difficult, as you will find if you try it for yourself! A hint makes it much easier. The conclusion of the message, . . . , is padding; you are told that the letter used for padding is } x. \text{ This gives a lot of information, since } . \text{ occurs twice in the rest of the message, and } x \text{ is usually preceded by } e \text{ in English; so we guess that } ^{\wedge} \text{ is } e. \text{ Now we have the sequence } e x \{ e ; ; \} \text{ which is probably going to be express, giving us three more letters. Now finish the rest yourself!}
\end{align*}
\]

The moral of this is that a seemingly innocent trait of the cryptographer, such as always using } x \text{ as a filler, may give away crucial information.

**Affine substitutions**

The sharp-eyed will have noticed something special about the substitution used here. It maps } a \text{ to } I, b \text{ to } D, c \text{ to } Y, \text{ and so on; advancing the plain letter one place moves the cipher letter back five places (or forward 21 places). In otherwords, if the letters}
of the alphabet are numbered from 0 to 25, so that \(a\) is represented by 0, \(b\) by 1, \ldots, \(z\) by 25, then the substitution takes the form

\[ x \mapsto 8 + 21x \pmod{26}. \]

Such a substitution, or the cipher it generates, is called **affine**.

The Caesar shift is a special case of an affine cipher, having the form

\[ x \mapsto x + b \pmod{26} \]

for some fixed \(b\). The general form of an affine cipher is

\[ x \mapsto ax + b \pmod{26} \]

for some fixed \(a\) and \(b\). The advantage is that the key is simple; instead of needing a general permutation of the letters, we only need the numbers \(a\) and \(b\) mod 26.

What affine ciphers are possible, and how can they be inverted?

First we must decide when an affine substitution is a permutation. Consider the substitution \(\theta : x \mapsto ax + b \pmod{n}\). It will fail to be a permutation if there exist two distinct \(x_1, x_2\) with

\[ ax_1 + b \equiv ax_2 + b \pmod{n}, \]

that is, \(ay \equiv 0 \pmod{n}\), where \(y = x_2 - x_1\). So \(\theta\) is a permutation if and only if the congruence \(ay \equiv 0 \pmod{n}\) has a solution \(y \not\equiv 0 \pmod{n}\).

Let \(d\) be the greatest common divisor of \(a\) and \(n\). Then, say, \(a = a_1d\) and \(n = n_1d\) for integers \(a_1, n_1\). Suppose that \(d > 1\), so that \(n_1 < n\). Putting \(y = n_1\), we have

\[ ay = a_1dn_1 = a_1n \equiv 0 \pmod{n}, \]

so \(\theta\) fails to be a permutation.

Conversely, suppose that \(d = \gcd(a, n) = 1\). By Euclid’s Algorithm (see the end of this chapter), there exist integers \(u, v\) such that \(au + nv = 1\). Now, if \(ay \equiv 0 \pmod{n}\), then

\[ y = (au + nv)y = u(ay) + n(vy) \equiv 0 \pmod{n}, \]

so \(\theta\) is a permutation.

We conclude:

**Theorem 1** The affine map \(x \mapsto ax + b\) is a permutation if and only if \(\gcd(a, n) = 1\).

What happens if we compose two such maps? Write \(\theta_{a,b}\) for the map \(x \mapsto ax + b\pmod{n}\), where \(\gcd(a, n) = 1\). We have

\[ \theta_{a,b} \circ \theta_{a',b'} : x \mapsto ax + b \mapsto a'(ax + b) + b', \]
so \( \theta_{a,b} \circ \theta_{a',b'} = \theta_{aa',ba'+b'} \).

The identity permutation \( x \mapsto x \) is the map \( \theta_{1,0} \). So to find the inverse of \( \theta_{a,b} \) in the form \( \theta_{a',b'} \), we have to solve the congruences

\[
\begin{align*}
  ad' &\equiv 1 \pmod{n}, \\
  bd' + b' &\equiv 0 \pmod{n}.
\end{align*}
\]

The first congruence has a unique solution mod \( n \), which can be found by Euclid’s Algorithm as before. Then the second congruence also has a unique solution, namely \( b' \equiv -bd' \pmod{n} \).

In particular, with \( n = 26 \), we want to invert the map \( \theta_{21,8} \). By trial and error (or by Euclid’s Algorithm), \( 21 \cdot 5 \equiv 1 \pmod{26} \); and then \( -5 \cdot 8 \equiv 12 \pmod{26} \). So the inverse of \( \theta_{21,8} \) is \( \theta_{5,12} \).

**Definition**  
*Euler’s totient function* \( \phi \) is the function on the natural numbers given by

\[
\phi(n) = \begin{cases} 
\text{number of congruence classes } a \bmod n \\
\text{such that } \gcd(a,n) = 1.
\end{cases}
\]

We give a formula for it, which will be proved later.

**Theorem 2**  
Let \( n = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r} \), where \( p_1, p_2, \ldots, p_r \) are distinct primes and \( a_1, a_2, \ldots, a_r \geq 0 \). Then

\[
\phi(n) = p_1^{a_1 - 1}(p_1 - 1)p_2^{a_2 - 1}(p_2 - 1) \cdots p_r^{a_r - 1}(p_r - 1).
\]

For example, \( 26 = 2 \cdot 13 \), so \( \phi(26) = 1 \cdot 12 = 12 \). The congruence classes coprime to 26 are represented by the odd numbers from 1 to 25 excluding 13.

**Theorem 3**  
The set of affine permutations mod \( n \) is a group of order \( n \cdot \phi(n) \).

We have verified the group properties in the earlier argument. For the order, note that there are \( \phi(n) \) choices for \( a \) and \( n \) choices for \( b \).

There are thus \( 26 \cdot 12 = 312 \) affine permutations. If we know or suspect that a substitution cipher is affine, we could try all 312 keys, though this is not trivial by hand. The method of looking for patterns of consecutive letters (as used to crack the Caesar cipher) does not apply. Like any substitution cipher, an affine cipher is vulnerable to frequency analysis. Its advantage is the small size of the key (two numbers rather than a complete permutation.)
**Worked example**  Decrypt the following affine substitution cipher:

JZQOU DQGKZ UULYU MKUOX LQJQJ ZQZCW ZQDYU MDXUJ
QRJCE LQEDR CRWGL UUIEJ JZQEP QDEWQ QEDRC RWGCR
JZCGK ZEDJJ ZQYJQ LLJZQ GJUDY

You are given that the frequency distribution in the ciphertext is as follows:

<table>
<thead>
<tr>
<th>C</th>
<th>D</th>
<th>E</th>
<th>G</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>O</th>
<th>P</th>
<th>Q</th>
<th>R</th>
<th>U</th>
<th>W</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>13</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>15</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>10</td>
</tr>
</tbody>
</table>

**Solution**  The commonest letter Q in the given cipher is likely to be e. We also see that the trigram JZQ occurs five times and so is likely to be the. This gives J=t and Z=h.

The letters Q and Z are \( x_{16} \) and \( x_{25} \) (where \( q = 26 \) here), while e and h and x_4 and x_7. Thus the parameters \( c \) and \( d \) satisfy

\[
4c + d \equiv 16 \pmod{26},
\]

\[
7c + d \equiv 25 \pmod{26},
\]

from which we find \( c = 3 \) and \( d = 4 \). Now we can compute the inverse of this affine transformation, which will be the decryption map: if the inverse is \( i \mapsto c'i + d' \), then we have (using the formula we worked out earlier)

\[
3c' = 1, \quad \text{so} \quad c' = 9;
\]

\[
4c' + d' = 0, \quad \text{so} \quad d' = 16.
\]

From this the entire substitution can be worked out, and we find the plaintext to be

themo resch oolyo ucomp letet hehig heryo urpot
entia learn insl ookat theav earge earni ngsin
thisc hartt heyte llthe story
or, correctly spaced and with punctuation,

The more school you complete, the higher your potential earnings. Look at the average earnings in this chart; they tell the story!

**Making a substitution cipher safer**

A substitution cipher can be solved by frequency analysis, and so is insecure for all but the shortest messages. However, there are some improvements that can be made. The first two rely on using a different alphabet for the ciphertext, with more characters than the plaintext alphabet. For example we could use an alphabet of 100 characters, represented by symbols 00,01,…,99.
Nulls: These are additional symbols in the cipher alphabet which do not have any meaning but are inserted in random positions to confuse the frequency analysis.

Homophones: We can translate the same letter in plaintext by several different letters in ciphertext. For example, if we use a 100-character cipher alphabet, we can associate about as many characters with each plaintext letter as its percentage frequency in normal text (say, 12 characters for e, 9 for t, and so on). Then we randomly decide which character to substitute for each occurrence of a letter. In the ciphertext, each character will occur approximately the same number of times. However, the ciphertext is still not random, and patterns of digraphs and trigraphs can be recognised.

Use of language: We can further confuse the analysis by using words from other languages, or by careful choice of words. As an example of what can be done, at least two English novels have been written containing no occurrence of the letter e, the commonest letter in English. One of these is Gadsby, by Ernest Vincent Wright. The author tied down the E key of his typewriter to write the book. The first paragraph reads as follows:

If youth, throughout all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically; you wouldn’t constantly run across folks today who claim that “a child don’t know anything.” A child’s brain starts functioning at birth; and has, amongst its many infant convolutions, thousands of dormant atoms, into which God has put a mystic possibility for noticing an adult’s act, and figuring out its purport.

To a casual glance, there is nothing odd about this; but it would pose an obvious problem for a cryptanalyst if encrypted with a substitution cipher. A frequency analysis of Gadsby is included in Table 1.

The novel A Void is even more remarkable, having been translated by Gilbert Adair from the French novel La Disparition by Georges Perec, which also lacked the letter e.

Another trick is to write words “phonetically”, or to use text-messaging abbreviations.

Features of text messaging language such as phonetic spelling (such as “nite” for “night”), the common omission of vowels (“txt” for “text”), use of abbreviations (such as AFAIK for “as far as I know”), use of numerals 2, 4 and 8 for to, for and ate, and use of “emoticons” such as ;-) as an essential part of the text, would give frequency analysis quite different from standard English. I don’t know whether such analysis of a body of text messages has been done.
Transposition: The substitution can be combined with transposition, that is, permuting the order of the characters in the ciphertext in a specified way. This will help to destroy the patterns of digram and trigram frequencies.

With these improvements, even a substitution cipher can be effective for a short message which will not receive very sophisticated analysis.

Number theory

In this section we give more details of some of the number theory which underlies our discussion of affine ciphers.

Euclid’s algorithm

Euclid’s algorithm is a procedure to find the greatest common divisor of two integers. In the form of a one-line recursive program it can be written as follows:

\[
\begin{align*}
\text{if } b &= 0 \text{ then } \text{gcd}(a,b) := a &\quad \text{else } \text{gcd}(a,b) := \text{gcd}(b, a \mod b) &\quad \text{fi}
\end{align*}
\]

where \( a \mod b \) means the remainder on dividing \( a \) by \( b \).

For example,

\[
\begin{align*}
\text{gcd}(30,8) &= \text{gcd}(8,6) = \text{gcd}(6,2) = \text{gcd}(2,0) = 2.
\end{align*}
\]

The algorithm can also be used to write \( \text{gcd}(a,b) \) in the form \( ua + vb \) for some integers \( u,v \). We express each successive remainder in this form until we reach the last non-zero remainder, which is the gcd. In the above example,

\[
\begin{align*}
6 &= 30 - 3 \cdot 8 \\
2 &= 8 - 1 \cdot 6 \\
&= 8 - (30 - 3 \cdot 8) \\
&= (-1) \cdot 30 + 4 \cdot 8,
\end{align*}
\]

so \( u = -1, v = 4 \).

This can be used to find inverses mod \( n \). For example, \( \text{gcd}(21,26) = 1 \), and Euclid’s algorithm shows that \( 1 = (-4) \cdot 26 + 5 \cdot 21 \); so \( 5 \cdot 21 \equiv 1 \pmod{26} \), and the inverse of 21 mod 26 is 5.
Euler’s function

In this section we prove Theorem 2. We begin with the theorem known as the Chinese Remainder Theorem.


There is an unknown number of objects. When counted in threes, the remainder is 2; when counted in fives, the remainder is 3; when counted in sevens, the remainder is 2. How many objects are there?

The problem asks for an integer \( N \) such that \( N \equiv 2 \pmod{3} \), \( N \equiv 3 \pmod{5} \), and \( N \equiv 2 \pmod{7} \). One solution is given as

\[
N = 2 \cdot 70 + 3 \cdot 21 + 2 \cdot 15 = 233;
\]

it is clear that adding or subtracting a multiple of 105 from any solution gives another solution; so the smallest solution is

\[
N = 233 - 2 \cdot 105 = 23.
\]

A folk-song gives the mnemonic:

Not in every third person is there one aged three score and ten,
On five plum trees only twenty-one boughs remain,
The seven learned men meet every fifteen days,
We get our answer by subtracting one hundred and five over and over again.

Why does it work? Observe that 70 is congruent to 1 mod 3, to 0 mod 5, and to 0 mod 7, and similarly for 21 and 15; then \( 70a + 21b + 15c \) is congruent to \( a \) mod 3, to \( b \) mod 5, and to \( c \) mod 7, as required. But how do we find these numbers 70, 21 and 15? Well, the first number is supposed to be divisible by 5 and 7, so is a multiple of 35; then 35 is congruent to 2 mod 3, so 2.35 is congruent to 2.2, which is congruent to 1 mod 3, as required. (In this last step we multiplied by the inverse of 2 modulo 3. In more difficult cases we can use Euclid’s algorithm to find the appropriate inverse.)

A similar procedure works in general. The fact that we can always find numbers with the required congruence conditions is not entirely obvious, but follows from Euclid’s algorithm using the fact that the moduli are coprime. We give the result just for two moduli: it is easily extended to any number by induction.

Let \( \mathbb{Z}/(n) \) denote the set of congruence classes mod \( n \). It is clear that, if \( x \equiv x' \pmod{mn} \), then \( x \equiv x' \pmod{m} \); so, for \( x \in \mathbb{Z}/(mn) \), there is a well-defined element \( x \mod m \) of \( \mathbb{Z}/(m) \). Similarly with \( n \) replacing \( m \).
Theorem 4 (Chinese Remainder Theorem) If $\gcd(m, n) = 1$, then the map $F$ from $\mathbb{Z}/(mn)$ to $\mathbb{Z}/(m) \times \mathbb{Z}/(n)$ defined by

$$F(x) = (x \mod m, x \mod n)$$

is a bijection.

Proof: Suppose that $F(x) = F(x')$. Then $x \mod m = x' \mod m$, that is, $m$ divides $x - x'$. Similarly $n$ divides $x - x'$. Since $m$ and $n$ are coprime, it follows that $mn$ divides $x - x'$, so that $x = x'$ (as elements of $\mathbb{Z}/(mn)$). Thus $F$ is one-to-one.

Now $|\mathbb{Z}/(mn)| = mn = |\mathbb{Z}/(m) \times \mathbb{Z}/(n)|$; so $F$ must also be onto, and thus a bijection.

This proof is non-constructive, whereas the original Chinese argument gave an algorithmic way to compute the inverse of $F$. This can be generalised as follows. Since $\gcd(m, n) = 1$, Euclid’s algorithm shows that there exist numbers $a$ and $b$ such that $am + bn = 1$. Now we see that

$$am \equiv 0 \pmod{m}, \quad am \equiv 1 \pmod{n},$$

$$bn \equiv 1 \pmod{m}, \quad bn \equiv 0 \pmod{n},$$

so the solution to the simultaneous congruences

$$x \equiv y \pmod{m}, \quad x \equiv z \pmod{n}$$

is given by

$$x \equiv bny + amz \pmod{mn}.$$ 

Remark: In fact $F$ is a ring isomorphism: this simply means that $F(x + x') = F(x) + F(x')$ and $F(xx') = F(x)F(x')$.

Now $\gcd(x, mn) = 1$ if and only if $\gcd(x, m) = 1$ and $\gcd(x, n) = 1$. Since Euler’s function gives the number of congruence classes coprime to the modulus, the Chinese Remainder Theorem implies that

$$\phi(mn) = \phi(m)\phi(n)$$

if $\gcd(m, n) = 1$.

This extends to products of any number of pairwise coprime factors. Thus

$$\phi(p_1^{a_1} \cdots p_r^{a_r}) = \phi(p_1^{a_1}) \cdots \phi(p_r^{a_r})$$

if $p_1, \ldots, p_r$ are distinct primes.

So, to complete the proof of the theorem, we have to show only that $\phi(p^a) = p^{a-1}(p - 1) = p^a - p^{a-1}$ for $p$ prime and $a > 0$. This holds because, of the $p^a$ congruence classes mod $p^a$, the ones not coprime to $p^a$ are precisely those divisible by $p$, of which there are $p^{a-1}$.