anyone cracking the codes. And there was more: Although encryption had been used for more than four millennia, no method known until the late twentieth century would have worked well enough for Internet commerce. But in 1976, two young mathematicians, operating outside the intelligence community that was the center of cryptography research, published a paper that made a reality out of a seemingly absurd scenario: Two parties work out a secret key that enables them to exchange messages securely—even if they have never met and all their messages to each other are in the open, for anyone to hear. With the invention of *public-key cryptography*, it became possible for every man, woman, and child to transmit credit card numbers to Amazon more securely than any general had been able to communicate military orders fifty years earlier, orders on which the fate of nations depended.

Historical Cryptography

Cryptography—"secret writing"—has been around almost as long as writing itself. Ciphers have been found in Egyptian hieroglyphics from as early as 2000 B.C. A *cipher* is a method for transforming a message into an obscured form, together with a way of undoing the transformation to recover the message. Suetonius, the biographer of the Caesars, describes Julius Caesar's use of a cipher in his letters to the orator Cicero, with whom he was planning and plotting in the dying days of the Roman Republic: "... if he [Caesar] had anything confidential to say, he wrote it in cipher, that is, by so changing the order of the letters of the alphabet, that not a word could be made out. If anyone wishes to decipher these, and get at their meaning, he must substitute the fourth letter of the alphabet, namely D, for A, and so with the others." In other words, Caesar used a letter-by-letter translation to encrypt his messages:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
DEFGHIJKLMNOPQRSTUVWXYZABC
```

To encrypt a message with Caesar's method, replace each letter in the top row by the corresponding letter in the bottom row. For example, the opening of Caesar's Commentaries "Gallia est omnis divisa in partes tres" would be encrypted as:

```
Plaintext:  GALLIA EST OMNIS DIVISA IN PARTES TRES
Ciphertext: JDOOLD HVW RPQLV GLYLVD LQ SDUWHV WUHV
```

The original message is called the *plaintext* and the encoded message is called the *ciphertext*. Messages are decrypted by doing the reverse substitutions.

This method is called the *Caesar shift* or the *Caesar cipher*. The encryption/decryption rule is easy to remember: "Shift the alphabet three places." Of course, the same idea would work if the alphabet were shifted more than three places, or fewer. The Caesar cipher is really a family of ciphers, with 25 possible variations, one for each different amount of shifting.

Caesar ciphers are very simple, and an enemy who knew that Caesar was simply shifting the plaintext could easily try all the 25 possible shifts of the alphabet to decrypt the message. But Caesar's method is a representative of a larger class of ciphers, called *substitution ciphers*, in which one symbol is substituted for another according to a uniform rule (the same letter is always translated the same way).

There are a great many more substitution ciphers than just shifts. For example, we could scramble the letters according to the rule

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ

XAPZRDWIBMQEOFTYCGSHULJVKN
```

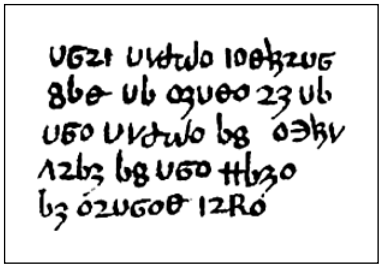
so that A becomes X, B becomes A, C becomes P, and so on. There is a similar substitution for every way of reordering the letters of the alphabet. The number of different reorderings is

$26 \times 25 \times 24 \times \cdots \times 3 \times 2$

which is about 4×10^{26} different methods—ten thousand times the number of stars in the universe! It would be impossible to try them all. General substitution ciphers must be secure—or so it might seem.

Breaking Substitution Ciphers

In about 1392, an English author—once thought to be the great English poet Geoffrey Chaucer, although that is now disputed—wrote a manual for use of an astronomical instrument. Parts of this manual, which was entitled *The Equatorie of the Planetis*, were written in a substitution cipher (see Figure 5.1). This puzzle is not as hard as it looks, even though there is very little ciphertext with which to work. We know it is written in English—Middle English, actually—but let's see how far we can get thinking of it as encrypted English.



Folio 30v of Peterson MS 75.1, *The Equatorie of Planetis*, a 14th century manuscript held at University of Cambridge.

FIGURE 5.1 Ciphertext in *The Equatorie of Planetis* (1392).

Although this looks like gibberish, it contains some patterns that may be clues. For example, certain symbols occur more frequently than others. There are twelve 6s and ten Us, and no other symbol occurs as frequently as these. In ordinary English texts, the two most frequently occurring letters are E and T, so a fair guess is that these two symbols correspond to these two letters. Figure 5.2 shows what happens if we assume that 6 = E and U = T. The pattern UG6 appears twice and apparently represents a three-letter word beginning with T and ending with E. It could be TIE or TOE, but THE seems more likely, so a reasonable assumption is that G = H. If that is true, what is the four-letter word at the beginning of the text, which begins with TH? Not THAT, because it ends with a new symbol, nor THEN, because the third letter is also new. Perhaps THIS. And there is a two-letter word beginning with T that appears twice in the second line—that must be TO. Filling in the equivalencies for H, I, S, and O yields Figure 5.3.

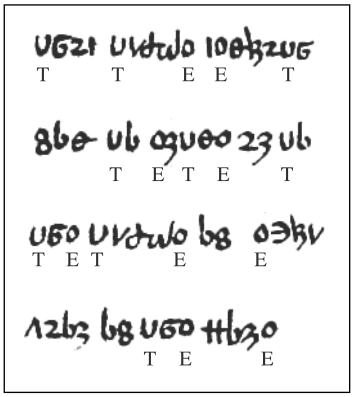


FIGURE 5.2 *Equatorie* ciphertext, with the two most common symbols assumed to stand for E and T.

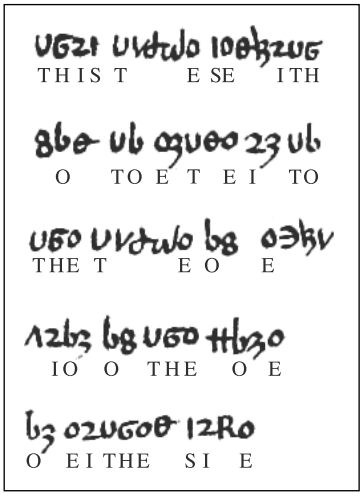


FIGURE 5.3 *Equatorie* ciphertext, with more conjectural decodings.

At this point, the guessing gets easier—probably the last two words are EITHER SIDE—and the last few symbols can be inferred with a knowledge of Middle English and some idea of what the text is about. The complete plaintext is: *This table servith for to entre in to the table of equacion of the mone on either side* (see Figure 5.4).

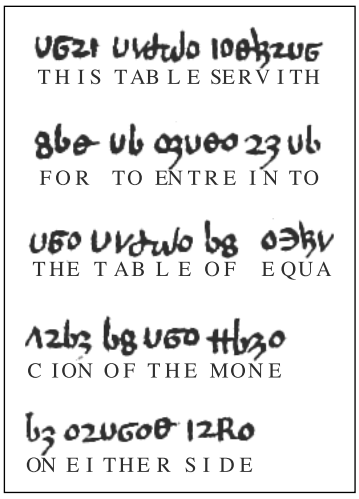


FIGURE 5.4 *Equatorie* ciphertext, fully decoded.

The technique used to crack the code is *frequency analysis*: If the cipher is a simple substitution of symbols for letters, then crucial information about which symbols represent which letters can be gathered from how often the various symbols appear in the ciphertext. This idea was first described by the Arabic philosopher and mathematician Al-Kindi, who lived in Baghdad in the ninth century.

By the Renaissance, this kind of informed guesswork had been reduced to a fine art that was well known to European governments. In a famous example of the insecurity of substitution ciphers, Mary Queen of Scots was beheaded in 1587 due to her misplaced reliance on a substitution cipher to conceal her correspondence with plotters against Queen Elizabeth I. She was not the last to have put too much confidence in an encryption scheme that looked hard to crack, but wasn't. Substitution ciphers were in common use as late as the 1800s, even though they had been insecure for a millennium by that time! Edgar Allen Poe's mystery story *The Gold Bug* (1843) and A. Conan Doyle's Sherlock Holmes mystery *Adventure of the Dancing Men* (1903) both turn on the decryption of substitution ciphers.

Secret Keys and One-Time Pads

In cryptography, every advance in code-breaking yields an innovation in code-making. Seeing how easily the *Equatorie* code was broken, what could we do to make it more secure, or *stronger*, as cryptographers would say? We might use more than one symbol to represent the same plaintext letter. A method named for the sixteenth-century French diplomat Blaise de Vigenère uses multiple Caesar ciphers. For example, we can pick twelve Caesar ciphers and use the first cipher for encrypting the 1st, 13th, and 25th letters of the plaintext; the second cipher for encrypting the 2nd, 14th, and 26th plaintext letters; and so on. Figure 5.5 shows such a Vigenère cipher. A plaintext message beginning SECURE... would be encrypted to produce the ciphertext *llqgrw*..., as indicated by the boxed characters in the figure—S is encrypted using the first row, E is encrypted using the second row, and so on. After we use the bottom row of the table, we start again at the top row, and repeat the process over and over.

We can use the cipher of Figure 5.5 without having to send our correspondent the entire table. Scanning down the first column spells out *thomasb-bryan*, which is the key for the message. To communicate using Vigenère encryption, the correspondents must first agree on a key. They then use the key to construct a substitution table for encrypting and decrypting messages.

When SECURE was encrypted as *llqgrw*, the two occurrences of E at the second and sixth positions in the plaintext were represented by different