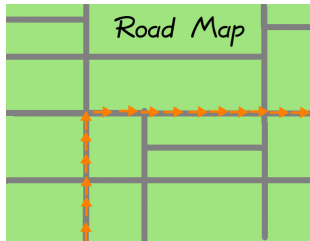


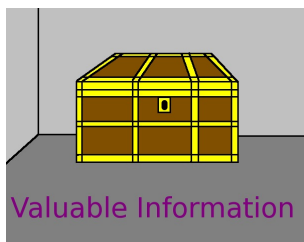
Module 10.1: Exploring Steganography with the Baconian Cipher



In this module, we'll explore the Baconian cipher. It is an example of *steganography*, which is the science of hiding one message inside another message—or in modern times, inside an image, an email, a sound file, an animation, a text message, or a movie. While it is from 1605, the Baconian cipher is very much related to modern problems in cybersecurity. Moreover, the Baconian cipher is very connected with binary numbers, which is not a concept that we normally associate with the early 17th century.

However, the Baconian cipher got its fame when theories were reported that Francis Bacon wrote some or all of the plays of William Shakespeare. Those theories were born because of the Baconian cipher, and they were not debunked until the 1950s.

A core part of the Baconian cipher (or I should say, the Baconian family of ciphers) is the following correspondence. Five symbol clusters, made up of As and Bs, are mapped to the English alphabet, in the following way, called the *Baconian alphabet table*.



AAAAA = A	ABAAA = I or J	BAAAA = R
AAAAB = B	ABAAB = K	BAAAB = S
AAABA = C	ABABA = L	BAABA = T
AAABB = D	ABABB = M	BAABB = U or V
AABAA = E	ABBAA = N	BABAA = W
AABAB = F	ABBAB = O	BABAB = X
AABBA = G	ABBBA = P	BABBA = Y
AABBB = H	ABBBB = Q	BABBB = Z

Now you might be wondering why I and J have the same symbol, and why U and V have the same symbol. The usual answer is that I and J are the same letter in Roman-era Latin, and were distinguished only in the medieval period. Likewise, U and V were the same letter in Roman-era Latin, and were only distinguished even later.

Indeed, this is why the letter W is called “double U” when, (as anyone can see) it clearly resembles a pair of Vs and not a pair of Us. Clearly, it should be called “double V,” but the name comes from an early period before U and V were distinct letters.

```
... 01001001 ...
... 00100000 ...
... 01001100 ...
... 01110101 ...
... 01110110 ...
... 00100000 ...
... 01000110 ...
... 01110011 ...
```

Look closely at the Baconian alphabet table. (That's the table given in the previous box.)

Clearly, this is binary code! Francis Bacon used A for 0 and B for 1, but you can clearly see that he starts from 00000, and counts in binary until 10111=23. It is exactly a binary counter, and this has several consequences.

First, you should never memorize the Baconian alphabet table, because you can always reproduce it when needed, effortlessly. In fact, you might want to write it out on a piece of scrap paper now. That's because you'll find that you will be frequently referring to the Baconian alphabet table throughout exploring this module.

Second, this really calls into question a lot of misconceptions that ordinary people have about the relative age of binary in particular, but also computer science and discrete mathematics in general.



It should be emphasized that Francis Bacon invented this cipher in 1605, and he had died in the year 1626. We tend to think of binary as a recent development, which is wrong. Moreover, all history-of-mathematics books that I've seen, at least as far as I can recall, attribute binary to Gottfried Wilhelm Leibniz (1646–1716). Francis Bacon is so much earlier that his lifetime does not even overlap the lifetime of Leibniz! We will discuss Francis Bacon later in this module, on Page 513.

You've probably heard of Leibniz because he is often described as the co-inventor of calculus, along with Isaac Newton (1642–1727). That's also false, or at least grotesquely over simplified, because Pierre de Fermat (1607–1665) knew about finding the optima of polynomial curves, and Archimedes (287 BCE—212 BCE) knew of some techniques that are startlingly similar to the integral calculus. Newton's teacher, Isaac Barrow (1630–1677) also had invented some bits of calculus, including the fundamental theorem of calculus.

Enough about calculus! Let's return to cryptography.

For Example :

Consider the following poem, which is clearly written in two different typefaces. As it turns out, a message has been hidden inside this poem, and the two fonts are crucial to recovering the message.

*TORCHES ARE MADE TO LIGHT, JEWELS TO WEAR,
DAINTIES TO TASTE, FRESH BEAUTY FOR THE USE,
HERBS FOR THEIR SMELL, AND SAPPY PLANTS...*

10-1-1

First, we will write a B under each fancy letter, and write an A under each normal letter.

<i>TORCHES</i>	ARE	MADÉ	TO	LIGHT,	JEWELS	TO	
BABBAAB	BAB	BAAB	BA	AABAA	AAAAAB	BA	
WEAR,	<i>DAINTIES</i>	TO	<i>TASTE,</i>	FRESH	BEAUTY	FOR	<i>THE</i>
AAAB	BBABAAAA	AA	BBAAB	AAAAA	AAABBA	ABA	BBA
USE,	<i>HERBS</i>	FOR	<i>THEIR</i>	SMELL,	AND	SAPPY	PLANTS...
ABA	BBAAB	AAB	AAABB	AAABA	AAA	AAABB	AAABAA...

Second, we will group the As and Bs by clusters of five. We get these clusters:

BABBA-ABBAB-BAABB-AAAABA-AAAAA-ABBAA-AABBB-ABAAA-AAAAB-ABAAA

AAAAA-ABBAA-BABBA-ABABB-AABAA-BAAAB-BAAAB-AAAAA-ABBA-ABAAA

Third, we use the 24-character Baconian alphabet table to convert the clusters of five As and Bs to letters of the ordinary English alphabet. In this case, we obtain the following:

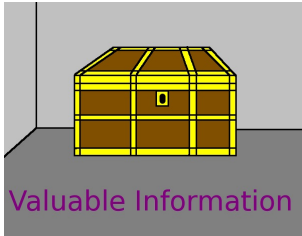
Y-O-U-C-A-N-H-I-D-E-A-N-Y-M-E-S-S-A-G-E

Therefore, we conclude that the message must be "You can hide any message."

If you are curious, the verses above are from *Venus and Adonis*, a poem published in 1593 by William Shakespeare (1564–1616). It is a good example of a poem that Francis Bacon might have chosen when he invented the Baconian cipher in 1605.

At that time, the word "biliteral" could be used to describe printing where two fonts have been used together, even inside the same word. For this reason, the Baconian cipher is sometimes called the biliteral cipher, and the Baconian alphabet table is sometimes called the biliteral alphabet. Bacon himself used this terminology in his lifetime, but it is no longer used today.

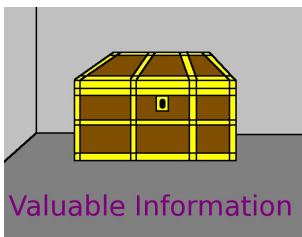
Some of my readers will already know the following technical vocabulary terms, but for completeness, I will define them here.



- The term *plaintext* represents the human-readable message.
- The term *ciphertext* represents the gibberish that you get after encryption.
- The process that takes a message from plaintext to ciphertext is called *encryption*. The sender *encrypts* a message so that it cannot be read if intercepted in transit.
- The process that takes a message from ciphertext to plaintext is called *decryption*, when done by the legitimate receiver.
- A pair of algorithms for encryption and decryption are together called a *cipher* or a *cryptosystem*. Modern computer scientists prefer to say “a cryptosystem” rather than “a cipher,” where as the term “cipher” is more often used when discussing techniques from the 1930s or earlier.
- Sometimes we will say that the receiver *decrypts* a message (usually for modern cryptosystems), and sometimes we will say that the receiver *deciphers* a message (usually for ciphers from the 1930s or earlier). Professors will also use the word “decipher” to sarcastically describe the tedious process of attempting to understand the writing of students with particularly bad handwriting.
- The cryptosystem that is, by far, the most often used on the internet is called *RSA*. The name comes from the first letters of the last names of the inventors: Ron Rivest, Adi Shamir, and Leonard Adleman. We will study RSA very thoroughly, throughout the later modules of this chapter.

Note: Sometimes a cryptosystem will include other algorithms, such as for setup. We will see an example of that when we learn about RSA. Since we will go into details later, we won't explore this point any further at this time.

Continuing with the list of vocabulary terms from the previous box, here are some more:

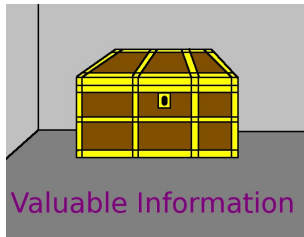


- *Steganography* is the science of hiding one message inside another. In hacker slang, this is called “steg,” as if to rhyme with “beg” or “keg.”
- The process that takes a message from ciphertext to plaintext, when done by some third party (not the legitimate receiver), is called *cryptanalysis*. A person who does this, particularly if they do it often, is called a *cryptanalyst*.
- The cryptanalyst is usually lacking significant amounts of information, which makes cryptanalysis vastly more difficult than decryption.
- In order for a cipher to be considered *secure*, cryptanalysis must be very nearly impossible in practice.
- The subject as a whole is called *cryptography* or *cryptology*. The distinction has now largely been forgotten. Long ago, cryptography would exclude cryptanalysis, while cryptology would include it. Both terms are in modern use, but cryptography is far more common.



Take care to make sure that you've gotten the terms in the previous box correctly internalized.

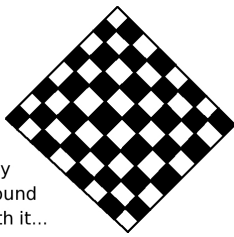
It is extremely common during technical interviews for jobs in computer science, computer engineering, or computer networking to pose an easy question that requires knowledge of these terms. A candidate who uses the vocabulary in a sloppy way will reveal themselves to be unfamiliar with cybersecurity, and that makes the candidate too risky to hire.



Generally, the steps for deciphering a Baconian message are as follows.

- Step 1: Convert the ciphertext to As and Bs.
- Step 2: Group the As and Bs into clusters of size five.
- Step 3: Substitute plaintext for As and Bs, using the Baconian alphabet table.

However, there is often a Step 0. If you're the intended recipient of the message, then you would know that the fancy letters were Bs and the normal letters were As. However, if you have intercepted the message, you do not necessarily know that. There must be a "Step 0: Figure out what the As and Bs are." In crypto competitions, that's often the crucial step and it can be difficult. We will explore this more as the module progresses.



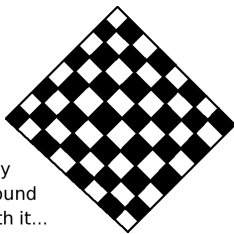
Play
Around
With it...

10-1-2

The following message was presented as the headline on a pamphlet, by John Toebes, about the National Science Olympiad "Codebusters" event at the Coach's Institute in the Summer of 2018.

BEING COVER AGENT FIXED DELAY PILOT RIGHT PLANE CATCH SMALL RADIO

Let the taller letters represent Bs, and the shorter letters represent As. What message do you get? The answer will be given on Page 519 of this module.



Play
Around
With it...

10-1-3

I'd like you to use the Baconian cipher to hide the message "Do not use large messages" inside the following snippet of a poem:

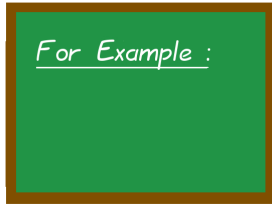
"... thou mean

To stifle beauty and to steal his breath,

Who when he liv'd, his breath and beauty set

Gloss on the rose, smell to the violet?"

Let the As indicate capital letters, and the Bs indicate lower-case letters, so that the poem will end up with a mix of capital and lower-case letters. (By the way, this is another snippet from *Venus and Adonis* by William Shakespeare.) The complete step-by-step solution is given on Page 519 of this module.



10-1-4

Imagine a very decorative botany book, such as a *pharmacopia*, where the margins have been decorated with some flowers. Some flowers face left, and some face right, but there is no obvious alternating pattern. Suppose they are grouped by threes, as follows.

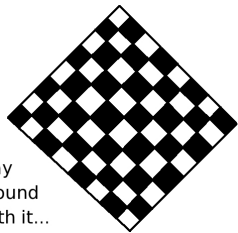
LLL-LRL-RRL-RRL-LRL-LLL-LLL-RRL-LLR-LLL-RLL-LRR-LLR-LRL-LLR-LRL-RLL-RLL-LLR-LLR-LLR-LLL-LRL-RLR-LRL-LRR-LRR-LRL-LLL-RLL-RLL-LLR-LLL-R

There are 33 clusters of 3, with one left over, so that's $(33 \times 3 + 1) = 100$ total Ls and Rs. This could be reorganized as 20 clusters of 5, and therefore we might suspect a Baconian cipher. Of course, we don't know if a left flower or a right flower represents the As.

Let's try R is A, and L is B, and let's see what happens. Grouping by five, we get this:
 BBBBA-BAABA-ABBAB-BBBBB-BAABB-BABBB-ABBBA-ABBAB-ABBBA-
 BABAB-BABBB-BABBA-BBABB-BBABA-BABAB-BAABA-ABABB-BBABB-ABBBB-
 ABBBA

Well, it looks like we're wrong, because we have some clusters of five that begin with BB. If you look at the Baconian alphabet table, you'll see that there are no letters of the alphabet associated with BB??? where the ? could represent either A or B. Those would, of course, represent the binary numbers for 24-31.

Next, we should try L=A and R=B. We'll continue in the following box.

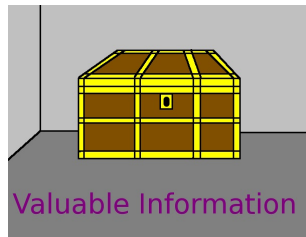


Play
Around
With it...

10-1-5

If we look at the Ls and Rs from the previous box, substitute L=A and R=B, what message do we get?

The answer will be given on Page 520.

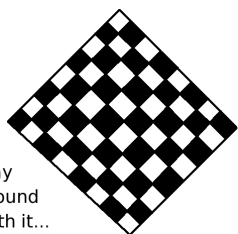


Valuable Information

Looking again at the previous example, there were two reasons that we should have known that R=A and L=B was wrong, but L=A and R=B was right. First, no cluster of five should begin with BB.

Second, the As are much more common than the Bs. For example, we have 34 Bs and 66 As in the correct version of the previous problem. Usually there are twice as many As as there are Bs, at least when the Baconian cipher is used with the English language. This fact can come in handy when working with much more complicated versions of the Baconian cipher. We will see some of those later in this module.

Whether we report it as 34 Bs, or 34% of the symbols are B, the technical term for that information is a *frequency count*. It is either the number of times a letter appears in either some plaintext or some ciphertext, or it can be stated in the form of a percentage.



Play
Around
With it...

10-1-6

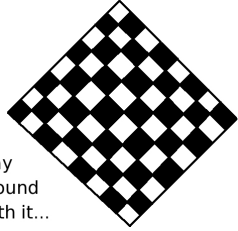
Here is a tweet from my friend John Toebes.



And yes that is Baconian!

8:10 AM - 7 Jan 2019

Can you decipher it? The solution will be given on Page 520.



Play
Around
With it...

10-1-7

Around the border of a page in a book about computer engineering, a student has found the following binary sequence of 18 bytes, plus one extra bit. Since that's $18 \times 8 + 1 = 145$ bits, a multiple of five, it seems likely that this might be a Baconian ciphertext. After all, it is somewhat unusual for the number of bits in a binary string to be a multiple of five.

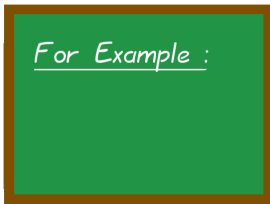
```
0100,0000.0001,0110.0000,0101.1000,0010.1110,0100.0001,1000.
0100,0000.0010,0110.1011,0001.0010,1100.0010,0101.0000,0000.
0001,0110.1100,1110.0100,0001.0100,0011.0000,0001.0000,1011.0
```

Permit me to save you some time, by allowing me to perform the frequency count for you. Among those 145 bits, there are 45 ones and 100 zeros. Based on what you learned in the previous box, I will allow you to decide whether you shall first try $A=0$ and $B=1$ or perhaps $A=1$ and $B=0$. Challenge yourself to recover the plaintext. (Remember, the plaintext is the human-readable message.)

The answer will be given on Page 520.

One of the easiest and most reliable cover stories for modern spies is to pretend to be a tourist. That way, there is every reason to wander around aimlessly and take photographs of just about everything, plus there is no expectation that you will speak the local language well. Of course, when a tourist acquires a travel visa for a large number of days, and the name on the application is one associated with a hostile power's intelligence agency, then the tourist might be closely monitored. Getting information out of the country, safely, might be very challenging, especially if one is closely monitored.

In the next box, we'll see a practical mechanism of doing just that, based on the Baconian cipher.



10-1-8

Suppose the arrangement for a spy (pretending to be a tourist) is for her to upload photographs to a common photo-sharing website like Instagram. She'll upload 0–15 photos everyday, and that will represent a number in binary, with the 0s being As and the 1s being Bs. This particular spy-pretending-to-be-a-tourist is driving around the Australian Outback, looking for a secret facility. What is more natural (and less suspicious) than a tourist uploading photographs to Instagram? She uploads the following counts of pictures over the first fifteen days of her trip:

0, 2, 9, 0, 2, 2, 4, 5, 13, 0, 4, 3, 0, 13, 1

First, we convert these numbers to binary:

```
0000-0010-1001-0000-0010-0010-0100-0101-1101-0000-0100-0011-0000-1101-0001
```

Second, we replace the 0s with As and the 1s with Bs. We'll do that in the next box.

Continuing with the previous box, we replace the 0s with As and the 1s with Bs, getting this:

```
AAAA-AAAB-BAAB-AAAA-AAAB-AAAB-ABAA-ABAB-BBAB-AAAA-ABAA-AABB-AAAA-BBAB-AAAB
```

Third, we regroup these into clusters of five. At this point, we have the following:

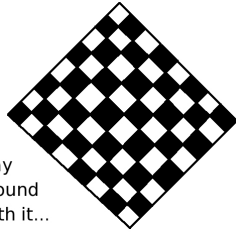
```
AAAAA-ABABA-ABAAA-AAABA-AABAA-BAAAB-ABBBA-BAAAA-ABAAA-ABBAA-AABBA-BAAAB
```

Fourth, we use the Baconian alphabet table to convert this into plaintext. We obtain what is below:

```
A-L-I-C-E-S-P-R-I-N-G-S
```

Therefore, our spy has determined (and is communicating back home) the location of the Australian/American base at Alice Springs.

If you'd like to learn more about the secret facility at Alice Springs, you can read the article "An American Spy Base Hidden in Australia's Outback," by Jackie Dent, published by *The New York Times* on November 23rd, 2017.



Play
Around
With it...

10-1-9

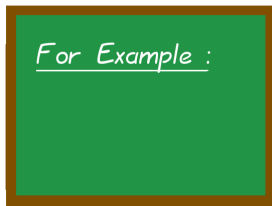
Another spy-pretending-to-be-a-tourist is going around unpopulated parts of the desert in the southwestern USA. She's using the same method as the previous example, but she'll upload 16 photographs to represent AAAA. This way, if there is a day where she forgets her charger (or cannot upload because of a busy schedule), the 0 can be recorded as a skipped day, rather than being recorded as

$$\text{zero} = 0000 = \text{AAAA}$$

In any case, here is the sequence of photograph-counts from the last ten days:

7, 16, 1, 4, 11, 1, 8, 1, 4, 4

What message does this represent? The answer will be given on Page 521 of this module.



10-1-10

The following message has only one font. That's useful because putting messages into two fonts alerts anyone who looks at the page that something is up, even if they don't know what is happening. A disadvantage of the following message is that it is still nonsensical gibberish, despite being in only one font.

A BOOK BE ITS A BOOK BACON EARTH A CASE BACK A A CALL
BE ITS A PLAN BACON ACRES ACRES BE ITS EARTH A CASE
BE ITS BACON A PLAN A HAND ACRES BE ITS EARTH A BOOK
ABBEY AMPLE A BOOK BACON A PLAN A CALL A CALL

As it turns out, the way that this message has been encoded is as follows. Half the alphabet encodes to A, using the mapping

$$\{A, C, E, G, I, K, M, O, Q, S, U, W, Y\} \rightarrow A$$

and the other half of the alphabet encodes to B, using the mapping

$$\{B, D, F, H, J, L, N, P, R, T, V, X, Z\} \rightarrow B$$

We will begin by converting the message into As and Bs, one word at a time. We will do that in the next box.

Continuing with the previous box, we first write out all the words of the previous message. If a letter comes from the A set, we will write an A under it. Of course, if a letter comes from the B set, we will write a B under it. We obtain this:

A	BOOK	BE	ITS	A	BOOK	BACON	EARTH	A	CASE	BACK
A	BAAA	BA	ABA	A	BAAA	BAAAAB	AABBB	A	AAAA	BAAA
A	A	CALL	BE	ITS	A	PLAN	BACON	ACRES	ACRES	BE
A	A	AABB	BA	ABA	A	BBAB	BAAAAB	AABAA	AABAA	BA
ITS	EARTH	A	CASE	BE	ITS	BACON	A	PLAN	A	HAND
ABA	AABBB	A	AAAA	BA	ABA	BAAAAB	A	BBAB	A	BABB
ACRES	BE	ITS	EARTH	A	BOOK	ABBEY	AMPLE	A	BOOK	BACON
AABAA	BA	ABA	AABBB	A	BAAA	ABBAA	AABBA	A	BAAA	BAAAAB
A	PLAN	A	CALL	A	CALL					
A	BBAB	A	AABB	A	AABB					

We will continue in the next box.

Continuing with the previous box, we group the As and Bs into clusters of fives, obtaining this:

```

ABAAA BAABA ABAAA BAAAB AABBB AAAAA BAAAA AAABB
BAABA ABBAB BAAAB AABAA AABAA BAABA AABBB AAAAA
BAABA BAAAB ABBAB ABABB AABAA BAABA AABBB ABAAA
ABBA AABBA ABAAA BAAAB ABBAB AAABB AAABB

```

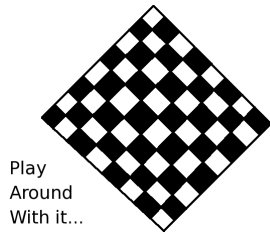
Third, we use the Baconian alphabet table to convert those clusters of five into English letters to get the plaintext:

I-T-I-S-H-A-R-D-T-O-S-E-E-T-H-A-T-S-O-M-E-T-H-I-N-G-I-S-O-D-D

Therefore, we can conclude that the message is “It is hard to see that something is odd.”

The following could pass as fairly low-quality post-modern poetry, of the type sometimes heard at late-night college poetry slams.

“Agile waltz!
Ardor in her;
beast as boy,
be its world.
As his dance ended,
do you award basil?
A king!
A game!
Dames as too awful death.
A baby,
a lady helps.”



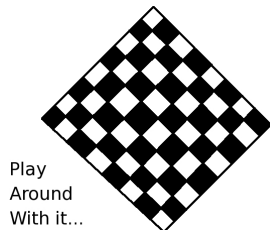
Play
Around
With it...

10-1-11

Yet, there is a message encoded there, using the Baconian cipher, similar to the previous example. Challenge yourself, and see if you can recover the message. The answer will be given on Page 521.

The following post-modern poetry is marginally better, but still fairly low-quality. It would be interesting to read it at a late-night college poetry slam, and see if anybody applauds.

“Apply a plan.
Agree, by any basic noble guide.
A kiss, if new.
A wish is now worth a case.
Do you court hairy Anton?
Bones, beams, a book, bombs is how!”



Play
Around
With it...

10-1-12

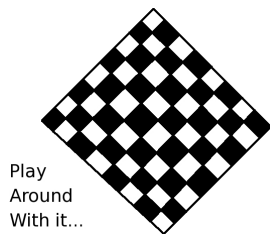
Challenge yourself, and see if you can recover the message. The answer will be given on Page 522.

Suppose an alien civilization gets their hands on a copy of this textbook, and they are so impressed with our knowledge, that they send the following message of good will to Earth. (Plus, this is a good way for them to verify if anyone has actually ever read this textbook.) Can you recover the plaintext?

BCFG-JLMP-RTUX-ZACF-HILN-PQTU-XZAC-FGJK-MPRT-VWYB-DFHJ-LNOQ-
TVXZ-BCEH-IKMP-QTVX-ZBCE-HIKN-PRSV-WYAD-FHIL-NORT-VX

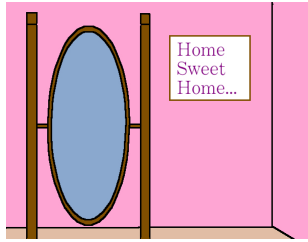
Hint: try to decrypt this one like you did with the previous two checkerboard boxes. That won't work. Look over what you've done, and figure out what needs to be changed. This is not as hard as it sounds.

The answer will be given on Page 522.



Play
Around
With it...

10-1-13

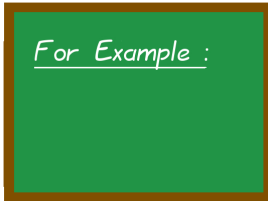


A Pause for Reflection...

The poems above aren't very good examples of slam poetry, and I shouldn't demean an expressive and popular genre of modern poetry. Here's a fine example of slam poetry where a college student recites his own poem. In this case, it is a slam poem that describes what it is like to live with depression.

<https://www.youtube.com/watch?v=GV-WONhEZec>

You might like to watch it, to give yourself a mid-module break while exploring this module's challenging topic. The video also provides insight into depression, which affects an enormous number of people world wide.



10-1-14

Sometimes, we have to face a little bit of uncertainty when working with a Baconian, if we don't know what should map to A and what should map to B. Consider the following ciphertext, where we are told to map the vowels to A and the consonants to B.

BAECI OUDFG YAHEI JOUYA EIKOU YLAEI MOUYN PAEIQ ORSUT YVAWX EIZOU YAEBC
 IODUY AEIOF UYAEI GOUHY AEJIO UKYAE IOLUM NYPQA ERIOU SYAET IOUYA VEIWX
 OZBUC DYFAE IOGUY AHEJI OKLUM NYAEI OPQUY ARSET VIOUWU

However, some linguists consider Y to be a vowel, and some consider Y to be a consonant, so we don't know how to map Y. That's okay, because instead of designating Y as an A or as a B, we will use a question mark, as you shall see. Let's carry out the mapping now—it is given in the next box.

Here is what happens when we take the ciphertext from the previous box replacing consonants with B, vowels with A, and Y with a question mark. We get the following:

BAECI	OUDFG	YAHEI	JOUYA	EIKOU	YLAEI	MOUYN	PAEIQ	ORSUT
BAABA	AABBB	?ABAA	BAA?A	AABAA	?BAAA	BAA?B	BAAAB	ABBAB
YVAWX	EIZOU	YAEBC	IODUY	AEIOF	UYAEI	GOUHY	AEJIO	UKYAE
?BABB	AABAA	?AABB	AABA?	AAAAB	A?AAA	BAAB?	AABAA	AB?AA
IOLUM	NYPQA	ERIOU	SYAET	IOUYA	VEIWX	OZBUC	DYFAE	IOGUY
AABAB	B?BBA	ABAAA	B?AAB	AAA?A	BAABB	ABBAB	B?BAA	AABA?
AHEJI	OKLUM	NYAEI	OPQUY	ARSET	VIOUWU			
ABABA	ABBAB	B?AAA	ABBA?	ABBAB	BAABA			

Many clusters of five have a ? in them. The very observant student will see that Y must be A, because if Y were B, we would have some clusters of five that begin with BB, and that's not supposed to happen. Now that we know Y=A, we can easily finish decoding the message.

However, suppose we didn't notice that right away. Can we handle uncertainty in a few locations? We'll see how to handle that, in the next box.

Continuing with the previous box, we have clusters of five, but with some spots unknown and shown with a question mark. We can use the Baconian alphabet table, and where a ? appears, we put both possibilities. Of course, if no ? appears in a cluster of five, we write only the one result. We get this:

BAABA T	AABBB H	?ABAA E or W	BAA?A R or T	AABAA E	?BAAA I/J or !?	BAA?B S or U	BAAAB S	ABBAB O
?BABBB M or !?	AABAA E	?AABB D or U/V	AABA? E or F	AAAAB B	A?AAA A or I/J	BAAB? T or U/V	AABAA E	AB?AA I/J or N
AABAB F	B?BBA Y or !?	ABAAA I/J	B?AAB S or !?	AAA?A A or C	BAABB U/V	ABBAB O	B?BAA W or !?	AABA? E or F
ABABA L	ABBAB O	B?AAA R or !?	ABBA? N or O	ABBAB O	BAABA T			

You can see that I have written !? when a five-letter cluster is called for, but does not exist in the Baconian alphabet table. (Those clusters begin with BB.) If the question mark is an A, meaning Y=A, there are no forbidden clusters, yet there are several forbidden clusters if Y=B. That reveals Y=A is to be used, meaning Y is considered a vowel here.

However, such “forbidden clusters” might not occur in other problems. Therefore, I should show you how to explore the problem if no “forbidden clusters” occur in either case.

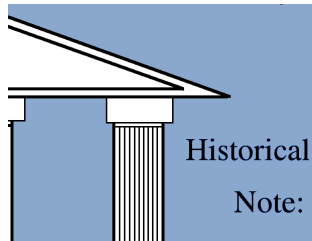
If we choose Y=B, meaning that Y is a consonant, we obtain this:

T-H-W-T-E-!?-U-S-O-!?-E-U/V-F-B-I/J-U/V-E-N-F-!?-I/J-!?-C-U/V-O-!?-F-L-O-!?-O-O-T

which looks like total gibberish. However, if we choose Y=A, meaning that Y is a vowel, we obtain this:

T-H-E-R-E-I/J-S-S-O-M-E-D-E-B-A-T-E-I/J-F-Y-I/J-S-A-U/V-O-W-E-L-O-R-N-O-T

and clearly, that’s the message. We have “There is some debate if Y is a vowel or not.”

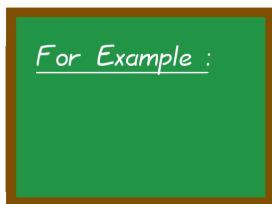


Historical
Note:

One of the major applications of the Baconian cipher in the twentieth century was the smuggling of messages out of totalitarian regimes by spies. For example, in many communist regimes, international mail was regularly opened and read by censors. If they saw anything suspicious, then the author could be in serious trouble. In World War II, even American servicemen (and women) had all of their letters home read, and censored with black ink.

There are times when you want to hide one message inside another, but unlike some of our previous examples, you don’t want it to be obvious that something is hidden. I’m now going to show an example of how to do that.

Consider the following message:



10-1-15

“I like to play the game of writing coded secrets. It built the mind and skill plus patience that honestly I find helped for learning to program in C and Java also. For sure, computer engineering needed even more raw, perfect patience from me often. Moreover, no career is free of boring moments. Growing patience is core for many jobs. Those with little patience are in need of some pain and math will give them lots. Little can be made in this life with few sacrifices of quality time or small loss of free hours. That’s just my perspective, truthfully.”

While it seems a bit unnatural at times, few people (who have not read this module) would suspect that the message above contains a hidden message inside it, protected with the Baconian cipher. However, it does have such a hidden message! Each word has its length counted. Then the odd-length words are Bs and the even-length words are As.

I 1 B	like 4 A	to 2 A	play 4 A	the 3 B	game 4 A	of 2 A	writing 7 B	coded 5 B	secrets. 7 B	S	H
It 2 A	built 5 B	the 3 B	mind 4 A	and 3 B	skill 5 B	plus 4 A	patience 8 A	that 4 A	honestly 8 A	O	R
I 1 B	find 4 A	helped 6 A	for 3 B	learning 8 A	to 2 A	program 7 B	in 2 A	C 1 B	and 3 B	T	M
Java 4 A	also. 4 A	For 3 B	sure, 4 A	computer 8 A	engineering 11 B	needed 6 A	even 4 A	more 4 A	raw, 3 B	E	S
perfect 7 B	patience 8 A	from 4 A	me 2 A	often. 5 B	Moreover, 8 A	no 2 A	career 6 A	is 2 A	free 4 A	S	A
of 2 A	boring 6 A	moments. 7 B	Growing 7 B	patience 8 A	is 2 A	core 4 A	for 3 B	many 4 A	jobs. 4 A	G	E
Those 5 B	with 4 A	little 6 A	patience 8 A	are 3 B	in 2 A	need 4 A	of 2 A	some 4 A	pain 4 A	S	A
and 3 B	math 4 A	will 4 A	give 4 A	them 4 A	lots. 4 A	Little 6 A	can 3 B	be 2 A	made 4 A	R	E
in 2 A	this 4 A	life 4 A	with 4 A	few 3 B	sacrifices 10 A	of 2 A	quality 7 B	time 4 A	or 2 A	B	E
small 5 B	loss 4 A	of 2 A	free 4 A	hours. 5 B	That's 5 B	just 4 A	my 2 A	perspective, 11 B	truthfully. 10 A	S	T

As you can see, the hidden message in the previous box was “Short messages are best.”

While that problem was too long and difficult for a test question, it is probably the safest way to actually use the Baconian in real life to convey a hidden message inside an email, a text message, or a tweet, if you ever need to do so.

Over the next three boxes, we’ll see an example where we make a hypothesis, and it turns out to be wrong. This happens from time to time in cryptography. A student must be valiant, and not give up at the first sign of trouble.

Consider the following message, encrypted with the Baconian cipher:

HFLMG-JHLKM-HFGLM-HJKFG-JKFLG-JMHKL-MFGHL-JKMFG-HJKFG-JLKFG-MJKFH
GJKFL-GMJHK-FLMGH-LJMKF-GHLJK-MFGJK-HFGLM-JHLKF-GMHJK-FLMGH-LJMKF

A quick frequency count reveals 17 Fs, 16 Gs, 15 Hs, 16 Js, 16 Ks, 15 Ls, and 15 Ms. Those frequencies are too close together (technical term: “too uniform”) to draw any conclusions of any kind.

Accordingly, we should look at the first two letters of each cluster of five, and see if we can use the fact that BB never occurs. There are 7 symbols here (F, G, H, J, K, L and M) so there are $C_{7,2} = 7(7-1)/2 = 7(6)/2 = 21$ pairings. (That’s the handshake principle, which we learned on Page 448, in the module “The Combinations and Handshake Principles.”)

We can see the following 2-letter prefixes among the 22 clusters of five:

HFx3, JHx2, HJx2, JKx2, JM, MFx2, JL, MJ, GJ, GMx2, FLx2, LJx2, GH

However, if we want to think of them as pairs, we should treat HJ and JH identically, JL and LJ identically, *et cetera*. We obtain this list:

Prefix Pairs seen: HF/FH x3, JH/HJ x4, JK/KJ x2, JM/MJ x2, MF/FM x2, JL/LJ x3, GJ/JG x1, GM/MG x2, FL/LF x2, GH/HG x1 (22 total, 10 distinct)

From there, it is easy to construct the list of pairs that were never seen, and that’s given below:

Prefix Pairs never seen: FG/GF, FJ/JF, FK/KF, GK/KG, GL/LG, HK/KH, HL/LH, HM/MH, KL/LK, KM/MK, LM/ML. (11 pairs never seen)

Also, we can be sure that we’ve found the complete set because we expect 21 pairs, 10 of which were seen, and 11 of which were never seen.

We will continue in the next box.

For Example :

10-1-16

Continuing with the previous box, at this point, we can reconstruct the set of symbols used for B, which we will denote \mathcal{B} . When I look at the end of the list of pairs never seen, I see that all possible pairings of K, L and M are never seen. So I’ll start with $\mathcal{B} = \{K, L, M\}$ and see if I can expand it. I should not add F to \mathcal{B} , because FL did occur. I should not add G to \mathcal{B} because GM is did occur. Yet, when I consider H, I realize that HK, HL, and HM never occurred. Thus, I should add H to \mathcal{B} getting $\mathcal{B} = \{H, K, L, M\}$. The only letter I haven’t considered is J, but JM did occur, so it cannot be in \mathcal{B} .

Of course, this means that $\mathcal{A} = \{F, G, J\}$. This is now enough information to recover the entire message. Substituting $\mathcal{B} = \{H, K, L, M\}$ with B, and $\mathcal{A} = \{F, G, J\}$ with A, we get this:

BABBA-ABBBB-BAABB-BABAA-ABABA-ABBBB-BAABB-ABBAA-BABAA-ABBAA-BABAB
AABAB-ABABB-ABBAB-BABBA-ABBAB-BAAAB-BAABB-ABBBA-ABBAB-ABBAB-BABBA

That translates to the following plaintext:

Y-Q-U/V-W-L-Q-U/V-N-W-N-X-F-M-O-Y-O-S-U/V-P-O-O-Y.

Sadly, that plaintext is clearly gibberish. Clearly, something has gone wrong. We’ll figure out what, in the next box.

Looking at the previous box, especially at the sequence of As and Bs above, it seems like there are too many Bs. We must have translated one extra symbol as B, because we have 49 As and 61 Bs after making our substitutions. Normally, we expect about twice as many As as Bs, and that clearly did not happen in this case.

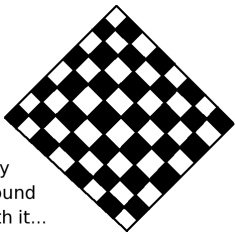
At this point, some work might be needed to figure out which symbol (or symbols) should be moved from \mathcal{B} to \mathcal{A} . To save time, permit me to tell you that K signifies A, not B. It is merely a coincidence that the prefixes KH/HK, KL/LK, and KM/MK never occur.

With this new information, we can make the revised substitutions: $\mathcal{A} = \{F, G, J, K\}$ for A, and $\mathcal{B} = \{H, L, M\}$ for B. After doing that, we obtain this:

BABBA-ABBAB-BAABB-BAAAA-AAABA-ABBAB-BAABB-AABAA-BAAAA-ABAAA-BAAAB
 AAAAB-ABABA-ABBAB-BABAA-ABBAA-BAAAA-BAABB-ABBAA-ABBAA-ABBAB-BABAA

That translates to the following plaintext:
 Y-O-U/V-R-C-O-U/V-E-R-I-S-B-L-O-W-N-R-U/V-N-N-O-W.

Therefore, we can conclude that the message is “Your cover is blown. Run now!”



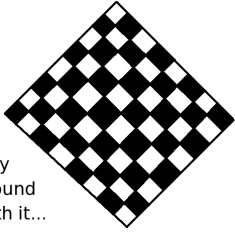
Play
Around
With it...

10-1-17

Suppose that the following cluster of numbers were to be found on a mathematician’s website. They are in clusters of five, so you might suspect that this is another example of the Baconian. Can you find the hidden message?

21346	58279	41635	82467	82946	18246	38254	68791	24368
25749	16824	36578	29468	12463	82468	24685	27941	36857
92416	83246	85724	96812	46382	46578	24968	12468	32465

Permit me to save you some time by giving you the frequency counts. Out of 135 digits, there are 10 ones, 22 twos, 10 threes, 22 fours, 10 fives, 22 sixes, 9 sevens, 21 eights, and 9 nines. The solution will be given on Page 523 of this module.



Play
Around
With it...

10-1-18

Suppose that this pattern was found at the bottom of the back cover of some game-design magazine.

$\uparrow \rightarrow \downarrow \quad \rightarrow \rightarrow \leftarrow \quad \uparrow \downarrow \leftarrow \quad \uparrow \downarrow \rightarrow \quad \leftarrow \uparrow \rightarrow \quad \downarrow \rightarrow \leftarrow \quad \uparrow \downarrow \leftarrow \quad \rightarrow \rightarrow \uparrow$
 $\downarrow \leftarrow \uparrow \quad \rightarrow \rightarrow \downarrow \quad \rightarrow \leftarrow \uparrow \quad \rightarrow \rightarrow \downarrow \quad \rightarrow \leftarrow \uparrow \quad \downarrow \leftarrow \uparrow \quad \downarrow \rightarrow \rightarrow \quad \leftarrow \uparrow \rightarrow$
 $\downarrow \leftarrow \uparrow \quad \rightarrow \rightarrow \downarrow \quad \rightarrow \leftarrow \uparrow \quad \rightarrow \rightarrow \downarrow \quad \leftarrow \uparrow \downarrow \quad \leftarrow \uparrow \downarrow \quad \rightarrow \leftarrow \rightarrow \quad \rightarrow \uparrow \downarrow$
 $\rightarrow \leftarrow \uparrow \quad \rightarrow \downarrow \leftarrow \quad \uparrow \rightarrow \downarrow \quad \rightarrow \leftarrow \uparrow \quad \downarrow \rightarrow \leftarrow \quad \uparrow \downarrow \rightarrow \quad \leftarrow \uparrow \rightarrow \quad \rightarrow \rightarrow \downarrow$
 $\leftarrow \uparrow \downarrow \quad \leftarrow \rightarrow \uparrow \quad \downarrow \leftarrow \uparrow \quad \downarrow \leftarrow \uparrow \quad \rightarrow \rightarrow$

There are 36 groups of three, plus two left over, and

$$(36)(3) + 2 = 108 + 2 = 110 = 22(5)$$

so one can suspect a Baconian cipher, since there is a multiple of five symbols. Can you recover the hidden message?

Permit me to save you some time. The frequencies are $\uparrow = 22.7272\%$; $\downarrow = 21.8181\%$; $\rightarrow = 33.6363\%$; and $\leftarrow = 21.8181\%$. The answer and full solution will be given on Page 524.

You might enjoy reading a description of the Baconian cipher published on April 25, 1925.
<https://toebes.com/Flynns/Flynns-19250425.htm>
 The linked article is from *Flynn’s Detective Fiction Weekly*, where a column “Solving Ciphers” appeared regularly, edited by Merle Ohaver. You’ll note that they use the older name, “the bilateral cipher,” instead of “Baconian cipher.” This column ran from December 13, 1924 until September 29, 1928, and led to the forming of the American Cryptogram Association (ACA) on September 1st, 1929.

As it turns out, there's some truly fascinating history involved with Francis Bacon and his cipher. The next twelve (!) boxes will explore that. As is the case with all the history boxes in this book, you can feel free to scroll past them if you don't enjoy reading about history.

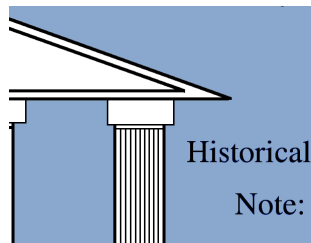


Few have influenced the world as much as Francis Bacon (1561–1626). In fact, it is very hard to write a summary without it growing to several pages in length. Francis Bacon was among the first to apply what would be later called scientific reasoning, and even extremely small amounts of statistics, to the day-to-day practice of running a major country.

To put this in perspective, Henry VIII (1491–1547), who had six wives and who had the heads of two of them cut off, had died only 14 years before Francis Bacon was born. Yet, by the time Francis Bacon had died, many modern institutions such as elections, newspapers, the modern idea of a nation's cabinet, patents for inventions, and the American colonies were well established.

While the scientific method has its roots in the Greek philosophers in general, and Aristotle of Athens (384 BCE–322 BCE) in particular, Francis Bacon is considered the father of modern empiricism because he revived this perspective, which had laid mostly ignored for centuries. Often, credit for this is given to Galileo (1564–1642) and René Descartes (1596–1650), who came a generation later.

We will explore the difference between Francis Bacon and physical scientists like Galileo and René Descartes in the next box.



René Descartes was interested in both spiritual questions (the existence of God and the nature of a soul), as well as practical matters, having made major discoveries in optics and also having invented the idea of coordinates in algebra and geometry. Galileo was interested in the physical world, having written books on astronomy, the strengths of materials, and especially kinematics—which is called “Physics 1” at most American colleges and universities. In contrast, Francis Bacon was interested in what would be called (centuries later) the social sciences—which were considered philosophy during his century and for several centuries afterward. However, it is better to consider Bacon a scientific philosopher than a scientist, because he did not make extensive use of mathematics or statistics.

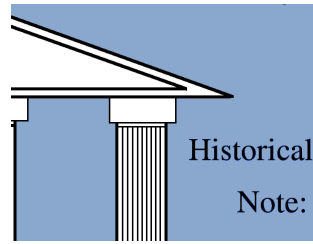
For example, there is an unverifiable story that he experimented by stuffing fowl with snow, to check a hypothesis that freezing meat would be a way of preserving it. (Of course, frozen meat would not become widely available for several centuries.) Unfortunately, he contracted pneumonia during this experiment and died.



Unlike most philosophers and early scientists of his period, Bacon did not at all confine himself to theory. Instead, he was very active in actually running England. He started out as a lawyer, then a member of parliament. He later became a judge, the Solicitor General of England, the Attorney General of England, and retired as the Lord Chancellor of England, a position that headed the whole judiciary but which also performed other functions, such as censorship.

Many claim that Bacon's rapid rise was due to his uncle, William Cecil (1520–1598), more often known as Lord Burghley, who was chief advisor to Queen Elizabeth I (1533–1603), and who held high positions such as Secretary of State and Lord Treasurer of England. The office of Prime Minister had not yet been invented, but Cecil basically played that role. However, Bacon's career isn't due to his uncle because Bacon's promotions didn't start until 1607, about nine years after his uncle had died.

This naturally brings up the question of what did help Francis Bacon. We will explore that in the next box.

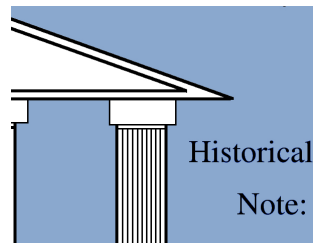


Historical
Note:

Several pivotal events in history were closely connected to Francis Bacon. Of course, we all know that Queen Elizabeth I was “the virgin Queen,” and (at least in theory) died a virgin. That’s why the US state named after Queen Elizabeth I is called Virginia and not Elizabethia. However, Queen Elizabeth had what we would now call “a crush” on a handsome military leader, Robert Devereux, better known as the Earl of Essex (1565–1601).

When the Earl of Essex led armed rebellion in 1601 against her, which was quickly crushed, Bacon helped lead, along with other lawyers, the prosecution of the Earl for treason. However, Francis Bacon was an experienced author. Once the Earl had been executed, Bacon wrote a pamphlet describing the events, the rebellion, the evidence, and the trial.

That pamphlet helped explain what had happened to English people, which was important because the Earl, as a dashing military hero, had been popular with many different layers of society.

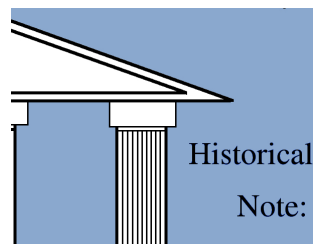


Historical
Note:

Being a virgin, Queen Elizabeth I had no children, and when she died in 1603, the throne of England passed to King James I (1566–1625), who was also King James VI of Scotland. This is sometimes written as King James VI/I. As it turns out, Queen Elizabeth’s paternal grandfather was King Henry VII (1457–1509), who was the father of the notorious King Henry VIII, and the great-grandfather of King James VI/I. In addition to the ancestral linkage, it probably helped that James and Elizabeth were both protestant and that James had been successful as King of Scotland.

Is it not amazing that Scotland and England, unified in 1603 after several centuries of being apart—including numerous wars—are to this very day united into one nation, more than 415 years later? To put this in perspective, the grandfather of King James VI/I was King James IV of Scotland (1473–1513), and he was killed in battle fighting England.

The English privy council (basically the equivalent of the modern cabinet) engaged in secret communications a full two years in advance of Queen Elizabeth’s death, to ensure a smooth transfer of the throne. The documents were kept secret from both parliament and the queen, but many of the documents are intact today. William Cecil, Francis Bacon’s uncle, was the primary point of contact on the English side, and it is known that they were using enciphered communications. Given the experience of Cecil’s nephew (Francis Bacon) with enciphered messages, it seems likely—but by no means certain—that Francis Bacon assisted in those communications and the peaceful transition which followed.



Historical
Note:

While Francis Bacon had openly supported James I prior to the transition, so had the Earl of Essex. To give you an idea of how close this was, the ambassadors sent by Scotland to England, for negotiating the transfer of the throne, left Scotland in 1601 before the Earl of Essex’s rebellion, intending to negotiate with him, but the rebellion was crushed and the Earl was executed before the ambassadors had arrived in London. Travel was slower in those days, and the Earl’s rebellion was pathetic and short-lived.

Bacon’s pamphlet, describing the rebellion and the Earl’s trial for treason, was an important factor in helping James I understand what had happened, preventing the new king from later taking revenge on the Earl of Essex’s opponents, and resulting in a (somewhat) politically unified England.

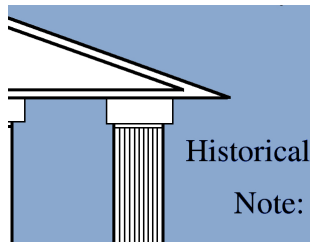
King James was himself a prolific author, and that common hobby of writing might also have helped Francis Bacon’s career. Bacon also helped negotiate some disputes between the House of Commons and the new king, but it would be tedious to discuss all the details.

The downfall of Francis Bacon is equally interesting, and we will now explore that in the next box.



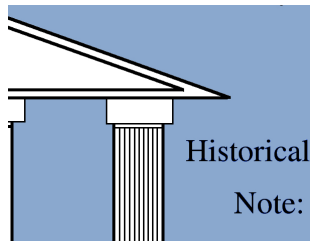
Unfortunately, in 1621 Francis Bacon was accused of having taken some bribes, for which he confessed. He was publicly disgraced and forced to retire. In fact, the House of Lords wanted him to be imprisoned for life, but King James ordered him released after only a few days.

Francis Bacon wrote profusely, and several books have been written about him, including about his sex life. He was 45 years old when he finally got married to Alice Barnham (1592–1650) who was 14 years old on her wedding day. However, he later disinherited her for adultery. Some theorize that Francis Bacon was mostly homosexual, which would explain his very late marriage, and they speculate that the reason he confessed so quickly to the bribery accusations was to prevent public accusations of buggery (anal sex between men), which was at the time a crime punishable by death.



To explain why the risk of buggery accusations was truly a credible threat, we will now discuss Francis Bacon's brother Anthony. Indeed, Anthony Bacon (1558–1601) had been publicly charged with buggery in 1586, thirty-five years earlier, when Anthony was working as a spy for England while living in Montauban, France. Henry IV (1553–1610), then King of Navarre but later also King of France, intervened and prevented Anthony from being punished. The penalty could have been being burnt at the stake.

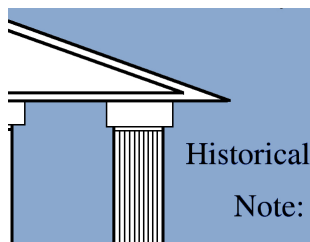
Actually, Anthony Bacon would escape a death sentence again 15 years later, as he was convicted for high treason as one of the supporters of the Earl of Essex and his 1601 rebellion. In fact, Francis Bacon played a role in his brother Anthony's prosecution! In the end, Anthony died before he could be executed.



We cannot be surprised that Francis Bacon invented a system for secret communications, given the major national and international intrigues that characterized his life and the life of his brother.

Francis Bacon was also heavily involved in theatre, which has led to some interesting theories that we will now explore. You have probably heard that there are some academic theories, as well as fringe theories, about the authorship of the plays of William Shakespeare (1564–1616).

We will explore one such theory, related both to Francis Bacon and the Baconian cipher, in the next box.



The earliest such theory was that a group of social reformers, including Francis Bacon and Walter Raleigh (1553±1–1618), for whom Raleigh, North Carolina is named, actually wrote all of them, so that William Shakespeare was just a conduit for publication and production of the plays. This theory was first put forward by Delia Salter Bacon (1811–1859), but has had many followers over the years, including until the present day.

As it comes to pass, Delia Salter Bacon claimed that the Baconian cipher was used to encode messages in the first printings of Shakespeare's plays. Indeed, the first editions sometimes used several different fonts mixed together, including adjacent letters of the same word. She had learned about ciphers from her friend, Samuel Morse (1791–1872), who is more famous for having invented Morse code.

We will address this connection between the Baconian cipher and Shakespeare's plays in the next box.

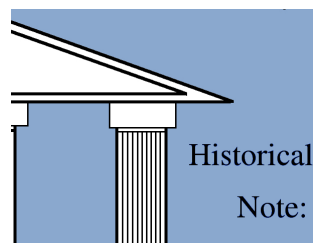


While cryptography cannot be used to address most questions in literature, two American cryptanalysts put forward a rigorous test of the hypothesis in the previous box. William Frederick Friedman (1891–1969) and Elizebeth Smith Friedman (1892–1980) debunked the theory of hidden messages in the first printings of Shakespeare’s plays. They were able to show, based on the statistical properties of the letter-font pairings in those first printings, as well as the common printing practices of the time, that there is no way that messages in the Baconian cipher are encoded in those first printings of Shakespeare’s plays. (Please note that Elizebeth Friedman is not a typo—she spelled her first name unusually, instead of the more commonplace Elizabeth.)

Of course, this does not address the question of whether or not Francis Bacon wrote Shakespeare’s plays, but it does address the question of whether or not secret messages were encoded in the first printings of Shakespeare’s plays using the Baconian Cipher. Moreover, a lot of the “evidence” quoted when advocating for Francis Bacon’s authorship involves enciphered messages.

In fact, the Friedmans wrote a book about all this, *The Shakespearean Ciphers Examined: An Analysis of Cryptographic Systems Used As Evidence That Some Author Other Than William Shakespeare Wrote the Plays Commonly Attributed to Him*, published by Cambridge University Press in 1957.

We will learn about the Friedmans in the next box.

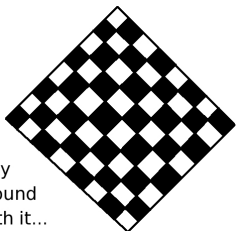


Interestingly, both Friedmans began their Shakespearean research during the first World War (1914–1919), with an eye toward proving that Bacon was indeed the author of Shakespeare’s plays, but the data lead them in the contrary direction.

The Friedmans built American governmental cryptanalysis from the first World War to the 1950s, and helped establish the NSA (National Security Agency). Moreover, the name of one or both Friedmans appears on many books and scholarly papers about cryptanalysis that were written in the middle of the twentieth century, but those are not well known outside of government circles because they were classified when written. One major cryptologic tool invented by William Friedman is “the index of coincidence,” used to break the Vigenère cipher, a major cipher in wide use from 1553 until the American Civil War (1861–1865). The Vigenère cipher was not invented by Vigenère however—an interesting story that will have to be told at some other time.

A major auditorium at NSA’s headquarters (in Fort Meade, Maryland) is named after the Friedmans, and it was still regularly used for training while I was working for the NSA during the years 1998–2002.

By the way, Aegean Park Press publicly reprinted several of William Friedman’s books once they had become declassified, which in some cases took several decades. You can easily find them for purchase on the internet.



Play
Around
With it...

10-1-19

Let’s return to the 21st century and talk about modern steganography now. To facilitate your understanding of the next box, please take a moment and convert the numbers 233, 124, and 89 into binary.

The solution can be found during the discussion inside the next box.

While steganography can occur anytime one message is inserted into another message, the most common application today on the internet is the hiding of a message inside of an image.

For example, with 24-bit color, and the RGB (red/green/blue) color scheme, there are 8 bits to define the shade of red, 8 bits to define the shade of green, and 8 bits to define the shade of yellow. Each color is represented by an integer from 0–255, which in binary is an 8-bit string. For example, 233 for red, 124 for green, and 89 for blue would be

```
... 01001001 ...
... 00100000 ...
... 01001100 ...
... 01110101 ...
... 01110110 ...
... 00100000 ...
... 01000110 ...
... 01110011 ...
```

11101001, 01111100, 01011001

in binary. There are $(2^8)(2^8)(2^8) = 16,777,216$ possible colors in this scheme. No human's eyes are that sensitive!

Realistically, you can flip the least significant bit, for each color, and no one would ever be able to notice. We would reconsider the previous color as

1110100?, 0111110?, 0101100?

where the question marks are bits that we can use for our hidden message. We would have either 232 or 233 for red, 124 or 125 for green, and 88 or 89 for blue. It is a difference of 1/256th of the way from “no red” to “maximum red,” and so forth for green and for blue.

This technique is called *least-significant bit steganography*, or *LSB-steganography*.

Continuing with the previous box, since those least-significant bits do not matter, you can take any image, and use the least significant bit (or maybe the two least significant bits) to encode your hidden message.

- You can use a Baconian alphabet table if you like, getting one letter for every five bits.
- You can use ASCII, getting one letter for every seven bits. This gives you access to numerals, symbols, and both capital and lower-case letters. If you are curious, ASCII stands for “American Standard Code for Information Interchange.”
- There is also “Half-ASCII” which has one letter for every six bits. While Half-ASCII includes all common punctuation and many symbols, it does not distinguish between capital and lower-case letters. Of course, neither does the Baconian alphabet table.
- The UTF-8 format includes ASCII when the most-significant bit is a zero. When the most-significant bit is a 1, you get even more symbols, useful for all modern languages, including non-alphabetic languages such as Chinese and Japanese Kanji, some ancient languages, and emoji. While ASCII characters require 8 bits, non-ASCII characters will require 16, 24, or 32 bits, depending on the character. If you are curious, UTF stands for “unicode transformation format.”
- Since it is very easy to search the internet for tables that provide the encodings for ASCII, Half-ASCII, and UTF-8, I will not provide them here.

These techniques work well for bitmapped images, which are not compressed. For more modern image-file formats (which are compressed) analogous procedures are possible, such as hiding information in the color table. The LSB-steganography concept also works well for sound files and with difficulty, movie files.

```
... 01001001 ...
... 00100000 ...
... 01001100 ...
... 01110101 ...
... 01110110 ...
... 00100000 ...
... 01000110 ...
... 01110011 ...
```

To give you an idea of how image steganography has been actually used by hackers over the past handful of years, you might want to read the following article, recommended by my advisee Kyle Conway. “Cybercriminals Use Malicious Memes that Communicate with Malware,” by Aliakbar Zahravi, published by *Trend Micro's Security Intelligence Blog* on December 14th, 2018.

```

... 01001001 ...
... 00100000 ...
... 01001100 ...
... 01110101 ...
... 01110110 ...
... 00100000 ...
... 01000110 ...
... 01110011 ...

```

Here is a note from another one of my paid proofreaders, Tanner Verber, about the LSB-steganography technique.

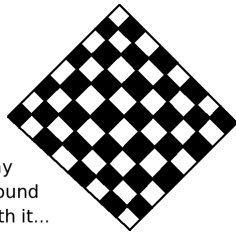
For my software engineering class, my group created a webpage that used the color hiding-steganography algorithm, and we were able to hide the entire works of Shakespeare in a picture of Shakespeare. This was a very fun project, ... (Major contributions were made by Evan van der Hoeven, Austin Scott, Connor Fergesen, Nicholas La Belle, and Brendan Bard.)

Can you imagine that?! They hid all the works of Shakespeare inside one image of Shakespeare! Moreover, a different paid proofreader, Trevor Kretschmann, also had a project about the LSB-steganography technique (in collaboration with others).

Let's take a moment to see if you've learned the vocabulary of this module correctly. Take the following 14 terms, and use them to fill in the blanks in the following sentences so that the sentences become true. Of course, since there are 15 spots and 14 vocabulary terms, at least one term will be used twice.

decrypt(s), frequency count(s), cryptology, cipher(s), decipher(s), encrypt(s), cryptosystem(s), decryption(s), plaintext(s), cryptanalysis, encryption(s), ciphertext(s), steganography, cryptography.

1. The unreadable gibberish that comes out of a good cryptosystem is called a _____.
2. The act of recovering the human-readable message by an interceptor, not the legitimate receiver is called _____.
3. The process that takes a message and makes it gibberish, so that no one can understand it if the message is intercepted, is called _____.
4. In the modern era, the receiver of a ciphertext must _____ it to make it readable. When speaking of classical ciphers from previous eras, we might say instead that the receiver _____ the ciphertext. The readable message is called the _____.
5. The science of hiding a message inside some other message is called _____.
6. The act of recovering the readable message by the legitimate receiver is called _____.
7. In order for a cipher to be considered secure, _____ must be very nearly impossible in practice.
8. The subject as a whole is usually called _____, but sometimes _____ instead.
9. The number of times a symbol appears in either some plaintext or some ciphertext, sometimes stated as a percentage, is called a _____.
10. The sender of a message _____ it, so that it cannot be read if intercepted in transit.
11. The more modern term for a pair of algorithms for encryption and decryption is a _____. The more classic term is a _____.



Play
Around
With it...

10-1-20

The answer will be given on Page 524 of this module.

This module is now complete. The remaining boxes have the solutions to earlier challenges.

Here is the solution to the question (from Page 503) about the headline written with capital letters of two different sizes. First, we're going to write an A under each smaller letter, and a B under each capital letter. When we do that, we get the following:

BEING	COVER	AGENT	FIXED	DELAY	PILOT
AAABA	ABBAB	AAABB	AABAA	AAAAB	BAABB

RIGHT	PLANE	CATCH	SMALL	RADIO
BAAAB	BAABA	AABAA	BAAAA	BAAAB

Next, we must use the Baconian alphabet table to convert these clusters of five into plaintext letters. We get this:

AAABA	ABBAB	AAABB	AABAA	AAAAB	BAABB
C	O	D	E	B	U/V

BAAAB	BAABA	AABAA	BAAAA	BAAAB
S	T	E	R	S

Clearly, the message is "Codebusters!"



Here I will show you how to embed the text "Do not use large messages" into that snippet of *Venus and Adonis*, as you were asked to do on Page 503.

First, we use the Baconian alphabet table to convert the plaintext message into clusters of As and Bs. Then, we get the following:

D	O	N	O	T	U	S
AAABB	ABBAB	ABBAA	ABBAB	BAABA	BAABB	BAAAB

E	L	A	R	G	E	M
AABAA	ABABA	AAAAA	BAAAA	AABBA	AABAA	ABABB

E	S	S	A	G	E	S
AABAA	BAAAB	BAAAB	AAAAA	AABBA	AABAA	BAAAB

We will continue in the next box.



Continuing with the previous box, we will use these As and Bs to make the letters of the poem capital and lower case. This means we will write the poem, in clusters of five (removing all spaces and punctuation) under the As and Bs. Remember, we were told to let the As indicate capital letters, and the Bs indicate lower-case letters. After doing that, we have this:

AAABB	ABBAB	ABBAA	ABBAB	BAABA	BAABB	BAAAB
THOum	EanTo	StiFL	EbeAu	tYAnD	tOSte	aLHIIs

AABAA	ABABA	AAAAA	BAAAA	AABBA	AABAA	ABABB
BReAT	HwHoW	HENHE	lIVDH	ISbrE	AThAN	DbEau

AABAA	BAAAB	BAAAB	AAAAA	AABBA	AABAA	BAAAB
TYsET	gLOSs	oNThe	ROSES	MEllT	OThEV	iOLEt

We will continue in the next box.

Continuing with the previous box, to make the plaintext human readable, we will put the spaces and punctuation back, following the pattern of the original poem. We get the following mess:

```
‘‘...THOU mEan
To StIFLE beAutY AnD tO SteaL HIS BReATH,
wHo WHEN HE LIV'D, HIS brEATH AND bEauTY sET
gLOSs oN The ROSE, SMELL TO The ViOLEt?’’
```

Optionally, we can make it look a heck of a lot better by changing the font, as you can see below:

```
“...THOU MEAN
TO STIFLE BEAUTY AND TO STEAL HIS BREATH,
WHO WHEN HE LIV'D, HIS BREATH AND BEAUTY SET
GLOSS ON THE ROSE, SMELL TO THE VIOLET?”
```



Here is the solution to the question about left-and-right facing flowers from Page 504. If we substitute L=A and R=B, and make clusters of five, we get the following:

```
AAAAB-ABBAB-BAABA-AAAAA-ABBAA-ABAAA-BAAAB-BAABA-BAAAB-
ABABA-ABAAA-ABAAB-AABAA-AABAB-ABABA-ABBAB-BABAA-AABAA-
BAAAA-BAAAB
```

We can look up the clusters of As and Bs in the Baconian alphabet table, and it translates into this:

```
B-O-T-A-N-I-S-T-S-L-I-K-E-F-L-O-W-E-R-S
```

Therefore, the message clearly is “Botanists like flowers.”



Here is the solution to the Baconian problem, from Page 504, which came from a tweet.

First, we notice that each horizontal line has exactly five “protrusions,” though some protrusions go up and some go down. Moreover, some protrusions are a single line, while other protrusions are a double line. Second, we see that only the middle protrusion of each horizontal line is a single—never double. Also, the other four protrusions of each horizontal line are doubles—never single. This implies that we can ignore the double/single distinction. Third, we have to figure out if up indicates A or B.

There are some horizontal lines that begin with down-down, up-down, and down-up, but none begin with up-up. Since BB cannot start a cluster of five in the Baconian alphabet table, this means that our first guess should be up=B and down=A. Fourth, we can use that to get clusters of five:

```
BABAA-ABAAA-BAABA-AABBB-ABBBA-BAAAA-AAAAA-AAABA-BAABA-ABAAA-AAABA-AABAA
```

Fifth, we can use the Baconian alphabet table to convert those clusters to English letters. We get W-I/J-T-H-P-R-A-C-T-I/J-C-E, so the plaintext must be “with practice.”



This is the solution to the question (from Page 505) about the 145-bit binary sequence. Regrouping the bits into clusters of fives, and then substituting 0=A and 1=B, we obtain the following:

```
ABAAA-AAAAA-ABABB-AAAAA-ABABB-AAAAA-BABBB-AABAA-AAABB-AAAAB-AAAAA
AAABA-ABBAB-ABBAA-ABAAB-ABBAA-AABAA-BABAA-AAAAA-AAAAB-ABBAB-BAABB
BAABA-AAAAB-ABAAA-ABBAA-AAAAA-BAAAA-BABBA
```

Next, using the Baconian alphabet table, we obtain this:

```
I-A-M-A-M-A-Z-E-D-B-A-C-O-N-K-N-E-W-A-B-O-U-T-B-I-N-A-R-Y.
```

Thus the message is clearly “I am amazed Bacon knew about binary.”

Here is the solution to the question (from Page 506) about a spy-pretending-to-be-a-tourist in the desert of the southwestern USA. We start with the photograph upload counts:

7, 16, 1, 4, 11, 1, 8, 1, 4, 4

First, we convert these numbers to binary:

0111-0000-0001-0100-1011-0001-1000-0001-0100-0100

Second, we replace the 0s with As and the 1s with Bs.

ABBB-AAAA-AAAB-ABAA-BABB-AAAB-BAAA-AAAB-ABAA-ABAA

Third, we regroup these into clusters of five. At this point, we have the following:

ABBBA-AAAA-ABABA-ABABB-AAABB-AAAA-ABABA-AABAA

Fourth, we use the Baconian alphabet table to convert this into plaintext. We obtain what is below:

P-A-L-M-D-A-L-E

Therefore, our spy has determined (and is communicating back home) the name Palmdale, California, which is the location of Skunkworks. Over several generations, many top-secret aircraft designs have been built and tested there.



Here is the solution to the question about post-modern poetry on Page 507. First, we convert the plaintext letters into As and Bs. Recall, we are using the mapping

$$\{A, C, E, G, I, K, M, O, Q, S, U, W, Y\} \rightarrow A$$

as well as the mapping

$$\{B, D, F, H, J, L, N, P, R, T, V, X, Z\} \rightarrow B$$

Substituting As and Bs gives us the following:

AGILE	WALTZ	ARDOR	IN	HER	BEAST	AS
AAABA	AABBB	ABBAB	AB	BAB	BAAAB	AA
BOY	BE	ITS	WORLD	AS	HIS	DANCE
BAA	BA	ABA	AABBB	AA	BAA	BABAA
ENDED	DO	YOU	AWARD	BASIL	A	KING
ABBAB	BA	AAA	AAABB	BAAAB	A	AABA
A	GAME	DAMES	AS	TOO	AWFUL	DEATH
A	AAAA	BAAAA	AA	BAA	AABAB	BAABB
A	BABY	A	LADY	HELPS		
A	BABA	A	BABA	BABBA		

We will continue in the next box.



Continuing with the previous box, we group the As and Bs into clusters of five. We get this:

AAABA AABBB ABBAB ABBAB BAAAB AABAA BAABA AABBB AABAA BABAA
 ABBAB BAAAA AAABB BAAAB AAABA AAAAA BAAAA AABAA AABAB BAABB
 ABABA ABABA BABBA

Third, we use the Baconian alphabet table to convert this to ordinary letters, getting the following:

C-H-O-O-S-E-T-H-E-W-O-R-D-S-C-A-R-E-F-U-L-L-Y

In conclusion, the message must be “Choose the words carefully.”

Here is the solution to the question about the second bit of post-modern poetry, given on Page 507.

First, we convert the plaintext into As and Bs.

APPLY	A	PLAN	AGREE	BY	ANY	BASIC	NOBLE
ABBBA	A	BBAB	AABAA	BA	ABA	BAAAA	BABBA
GUIDE	A	KISS	IF	NEW	A	WISH	IS
AAABA	A	AAAA	AB	BAA	A	AAAB	AA
NOW	WORTH	A	CASE	DO	YOU	COURT	HAIRY
BAA	AABBB	A	AAAA	BA	AAA	AAABB	BAABA
ANTON	BONES	BEAMS	A	BOOK	BOMBS	IS	HOW
ABBAB	BABAA	BAAAA	A	BAAA	BAABA	AA	BAA

We will continue in the next box.



Continuing with the previous box, second, we group the As and Bs into clusters of five. Third, we use the Baconian alphabet table to convert the clusters of five into plaintext letters.

ABBBA	ABBAB	AABAA	BAABA	BAAAA	BABBA	AAABA	AAAAA
P	O	E	T	R	Y	C	A
ABBAA	AAAAB	AABAA	AABBB	AAAAA	BAAAA	AAABB	BAABA
N	B	E	H	A	R	D	T
ABBAB	BABAA	BAAAA	ABAAA	BAABA	AABAA		
O	W	R	I/J	T	E		

Finally, we conclude that the message must be “Poetry can be hard to write.”

Here is the solution to the problem (from Page 507) where we are trying to decipher some communications from aliens who have read this textbook.

BCFG	JLMP	RTUX	ZACF	HILN	PQTU	XZAC	FGJK
BABA	BBAB	BBAB	BAAB	BABB	BABA	BBAA	BABA
MPRT	VWYB	DFHJ	LNOQ	TVXZ	BCEH	IKMP	QTVX
ABBB	BAAB	BBBB	BBAA	BBBB	BAAB	AAAB	ABBB
ZBCE	HIKN	PRSV	WYAD	FHIL	NORT	VX	
BBAA	BAAB	BBAB	AAAB	BBAB	BABB	BB	



The brightest students will immediately have observed that there are simply too many Bs. Nonetheless, let’s proceed by regrouping these into clusters of five. We get

BABAB-BABBB-ABBAA-BBABB-BABAB-BAABA-BAABB-BBAAB-BBBBB
 BAABB-BBBAA-BAAAB-ABBBB-BAABA-ABBBA-BAAAB-BBABB-ABBBB

which is problematic, since 4 of the clusters start with BB, and that’s forbidden.

We might hypothesize that the aliens have reversed A and B. We could check this by noting that there is no cluster that starts with AA. That means after flipping the As and Bs, there would be no cluster with BB.

Another approach entirely is to count the As and Bs, getting 34 As and 56 Bs among the 90 letters. Normally we’d expect 60 As and 30 Bs, so that’s further evidence that we should try flipping them. We’ll do that in the next box.

Continuing with the previous box, after flipping, we obtain the following.

ABABA	ABAAA	BAABB	AABAA	ABABA	ABBAB	ABBAA	AABBA	AAAAA
L	I	V	E	L	O	N	G	A
ABBAA	AAABB	ABBBA	BAAAA	ABBAB	BAAAB	ABBBA	AABAA	BAAAA
N	D	P	R	O	S	P	E	R

The message is clearly “Live long, and prosper!” By the way, this message has unusually many letters from the later part of the alphabet, but only three letters before “i.” That’s slightly unusual, and it explains why we have 56 As and 34 Bs, instead of something closer to 60 As and 30 Bs.



Here is the solution to the question (from Page 512) about the clusters of five integers. The frequency count told us that even numbers are twice as common as the odd the numbers. Therefore, we might have the hypothesis that the set $\{2, 4, 6, 8\}$ represent the As and that the set $\{1, 3, 5, 7, 9\}$ represent the Bs. However, it would be wise for us to double-check that hypothesis before investing further time.

The way to do that is to search for any clusters of five that start with two odd numbers. Under our hypothesis, that would result in a cluster beginning with BB, which is forbidden. A quick scan of the clusters reveals that there are no clusters beginning with two odd numbers. Therefore, we can proceed with more confidence.

We can also double check by totaling the frequency counts of the even numbers and comparing them to the odd numbers. We get 48 odd numbers and 87 even numbers, which seems reasonable. (To be precise, we observed 35.5555% odd and 64.4444% compared with 33.3333% and 66.6666% being expected).

We will continue in the next box.

Continuing with the previous box, we should now replace the even numbers with A and the odd numbers with B. We obtain the following:

21346	58279	41635	82467	82946	18246	38254	68791	24368
ABBAA	BAABB	ABABB	AAAAB	AABAA	BAAAA	BAABA	AABBB	AABAA
N	U	M	B	E	R	T	H	E
25749	16824	36578	29468	12463	82468	24685	27941	36857
ABBAB	BAAAA	BABBA	ABAAA	BAAAB	AAAAA	AAAAB	ABBAB	BAABB
O	R	Y	I	S	A	B	O	U
92416	83246	85724	96812	46382	46578	24968	12468	32465
BAABA	ABAAA	ABBAA	BAABA	AABAA	AABBA	AABAA	BAAAA	BAAAB
T	I	N	T	E	G	E	R	S

The message clearly is “Number theory is about integers.”

Now for the solution to the question with the groups of arrows in the game-design magazine, from Page 512. We first have to figure out which symbols are As and which symbols are Bs. Let's look at the frequencies again. The frequencies are 22.7272% \uparrow , 21.8181% \downarrow , 33.6363% \rightarrow , and 21.8181% \leftarrow .

In theory, there could be multiple symbols for B. However, in this case, that seems very unlikely. If we add the two rarest symbols together, we get $2(21.8181\%) = 43.6362\%$ which is too high. We expect about $1/3$ of the symbols to be B. Thus there is probably only one symbol that is B. Moreover, the frequency of \rightarrow is suspiciously close to $1/3$, being 33.6363% in place of 33.3333%. Therefore, we should hypothesize that \rightarrow is B, and the other symbols are A. To check our work, we can add the frequencies of the As, getting

$$22.7272\% + 21.8181\% + 21.8181\% = 66.3634\%$$

which is very nearly $2/3$. Our hypothesis seems good, at least so far.

Next we convert the \rightarrow s to B, and the other arrows to A. After doing that, and grouping by fives, we get the following:

ABABB	AAAAA	ABAAB	ABAAA	ABBAA	AABBA	BAABB	ABAAA
M	A	K	I	N	G	V	I
AAABB	AABAA	ABBAB	AABBA	AAAAA	ABABB	AABAA	BAAAB
D	E	O	G	A	M	E	S
	ABAAA	BAAAB	AABBB	AAAAA	BAAAA	AAABB	
	I	S	H	A	R	D	

Therefore, the message is "Making video games is hard."



Here is the solution to the vocabulary-matching question from Page 518.

1. The unreadable gibberish that comes out of a good cryptosystem is called a ciphertext.
2. The act of recovering the human-readable message by an interceptor, not the legitimate receiver is called cryptanalysis.
3. The process that takes a message and makes it gibberish, so that no one can understand it if the message is intercepted, is called encryption.
4. In the modern era, the receiver of a ciphertext must decrypt it to make it readable. When speaking of classical ciphers from previous eras, we might say instead that the receiver deciphers the ciphertext. The readable message is called a plaintext.
5. The science of hiding a message inside some other message is called steganography.
6. The act of recovering the readable message by the legitimate receiver is called decryption.
7. In order for a cipher to be considered secure, cryptanalysis must be very nearly impossible in practice.
8. The subject as a whole is usually called cryptography, but sometimes cryptology instead. (Don't worry if you got those two terms reversed, it is fine.)
9. The number of times a symbol appears in either some plaintext or some ciphertext, sometimes stated as a percentage, is called a frequency count.
10. The sender of a message encrypts it, so that it cannot be read if intercepted in transit.
11. The more modern term for a pair of algorithms for encryption and decryption is a cryptosystem. The more classic term is a cipher.

