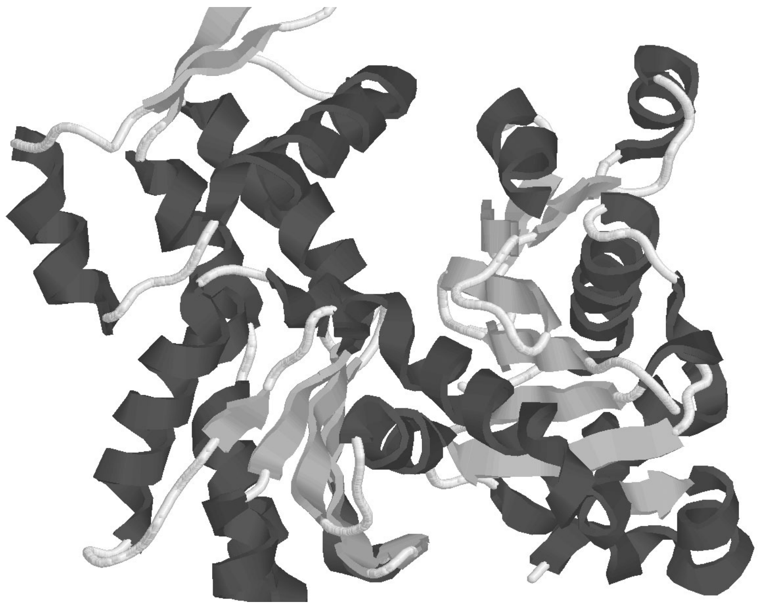
Part 1: The Evolution of Knowledge and Information



Cartoon representation of the crystal structure of the protein, actin. Actin is one of the proteins required for muscle contraction.

Chapter 1: Information vs. Knowledge

One of the goals in this book is to investigate how molecular knowledge evolves in biological systems. Because there is no mathematical definition for knowledge, much of this investigation will be open to interpretation. To find the molecular knowledge possessed by a chemical like DNA, it is first necessary to quantify the amount of information. This is true because unlike knowledge, information has a precise mathematical definition.

The scientific definition for information is very different from the common one. The everyday definition implies that information should be useful or at least convey some amount of knowledge. The scientific definition does not make any distinction between useful and useless information. For example, consider the following sentences:

The brown dog likes to fetch a tennis ball.

Zxrd zgbzbue awfllt jhjzhwzhg zwnzi oppwnnni wyxaz.

If information theory is applied to these two sentences, the results will indicate that the second sentence contains much more information than the first. Not only is the second sentence longer, but it contains many letters that are rarely used in English (z, x, and w). The first sentence contains useful information. The second message contains no useful information. Yet information theory asserts that the second contains more information than the first. How can this be?

To understand why, consider why scientists developed information theory. The theory was developed by an engineer, Claude Shannon, who was interesting in transmitting information. The second sentence takes longer to transmit than the first, so it contains more information. This definition is clearly not useful to biologists studying evolution.

Evolution involves the creation of information that provides a selective advantage. That is the organism that possesses the new information has an edge over those that do not. Therefore, the information must be useful. This is why the word knowledge is preferable. Knowledge implies that information is useful.

In communication systems, information does not have to contain knowledge. In general, the same cannot be said for biological systems. Information that does not provide a selective advantage is often lost. Thus, the information found in biological systems usually conveys knowledge, and this knowledge provides a selective advantage. In biological systems, knowledge and information are often related, but they are not necessarily equal. Consider the following two sentences:

I have a dog. His name is Bubba. He is a black lab. He is 13 years old.

My black lab, Bubba, is 13.

Both sentences describe four identical concepts, so the knowledge conveyed by both is identical, but the first sentence contains much more information than the second. Because information has a precise mathematical definition, it can be determined rather easily. In contrast, knowledge will always be open to interpretation. Nevertheless, it is possible to define molecular knowledge in terms of information. The proposed definition is as follows:

Molecular Knowledge: the minimum amount of information necessary to enable a chemical (or group of chemicals) to accomplish some task or to specify some trait. The only stringent requirement is that molecular knowledge must confer a selective advantage so that natural selection can preserve it.

Because molecular knowledge is now defined in terms of information, information theory can be used in conjunction with human insight to calculate knowledge. The rest of this chapter will explore information and its properties.

The Nature of Information

Mathematically, information is defined as a reduction in uncertainty. Consider a scientist trapped in a room. He has a coin and a telephone. He is told to flip the coin and then tell his colleagues who are 500 miles away the results using the telephone. He is to repeat this process until he is told otherwise.

Before the scientist flips the coin, he does not know whether it will land heads or tails. There are two possible outcomes, and the scientist does not know which will happen until he observes the results. Suppose that the first toss is heads. As soon as the scientist observes this result, he has information. Two possibilities have been reduced to one. Before observing the coin, the scientist was uncertain of the outcome. After he observes the result, he is certain of the outcome. His colleagues do not have any information until he tells them that the coin landed heads.

A unit of information is called a bit. Whenever two possible outcomes are reduced to one, one bit of information is created. Thus, the scientist acquires one bit of information each time he tosses the coin and observes the result.

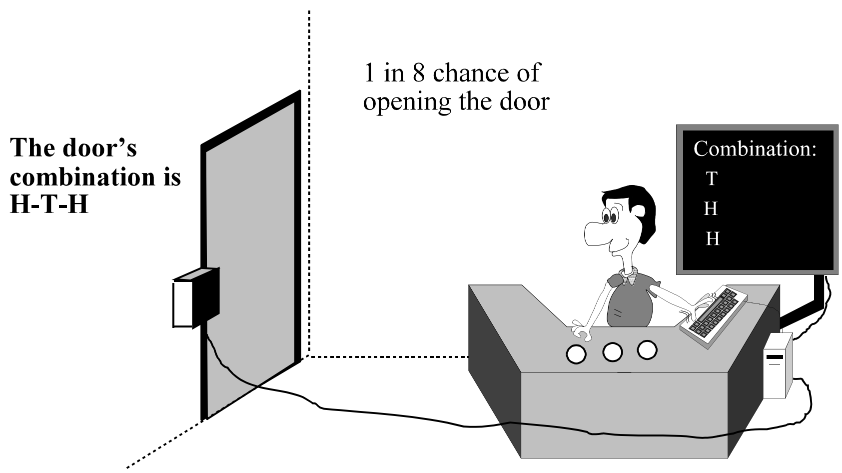
Uncertainty depends on the number of possible outcomes. For example, if the scientist is given a die to roll instead of a coin his uncertainty increases. With the coin, there are only two possible outcomes, and with a die there are six. The reduction in uncertainty for a die (six possible outcomes reduced to one outcome) is greater than it is for the coin (two possible outcomes reduced to one outcome).

Thus, the scientist acquires more information when he observes the results of tossing the die than when he observes the results of tossing the coin.

Trapped Scientist with Three Coins

Consider a scientist trapped in a room. He is given three coins, instructed to toss the coins one at a time, and enter the results into the computer (figure 1.1). If the coin lands tails, he is instructed to enter the letter *T*, and if it lands heads, he is instructed to enter the letter *H*. The combination for the door is H-T-H. The scientist has a 1 in 8 chance of opening the door.

Figure 1.1: Trapped Scientist with Three Coins



When three coins are tossed, there are eight possible outcomes:

Coin 1	Coin 2	Coin 3	
H	H	H	
H	H	T	
H	**T**	**H**	--------->opens the door
H	T	T	
T	H	H	
T	H	T	
T	T	H	
T	T	T	

When the first coin is tossed, it has two ways to land, heads or tails. The same rules apply for the second and third. So the total number of possible outcomes is 2x2x2 = 8.

Whenever the scientist tosses a coin and observes the results, he acquires information. When he tosses the first coin and observes the result, he acquires one bit of information. After he observes the result of the second coin, he possesses 2 bits of information, and after the third, he possesses 3 bits. Suppose on his first try to open the door, all three coins land heads. After observing this event, the scientist will possess 3 bits of information. He keeps trying, and after a few more tries, the first coin lands head, the second lands tails and the third lands heads. When he enters this result into the computer, the door opens. The scientist has acquired knowledge. The combination for the door is H-T-H, and he now knows the combination.

Notice that every time the scientist tosses the coin he creates information, but only one specific outcome creates useful information or knowledge.

One bit of information corresponds to each coin. In figure 1.1, all results contain 3 bits of information. One result, H-T-H, contains 3 bits of knowledge.

Suppose that the combination is changed to H-T-H-H-H-H-T-H-H-H-H-H-H-H-H-H-H-H-H-H. The scientist is given 20 coins, told to toss all 20, enter the results into the computer and observe the door. How much information is generated every time the scientist tosses 20 coins and observes the result? Answer: 20 bits because there are 20 coins. While 20 bits of information is generated with each attempt to open the door, only one possible outcome will open the door. This is the only outcome that contains both information and knowledge.

With 20 coins, what is the probability that the scientist will find the correct combination on the first try? Answer: multiply 2 by itself 20 times to determine the total number of possible outcomes (1,048,576). Because only one of these outcomes will open the door, the odds are 1 in 1,048,576 or approximately 1 in a million.

Exponents are a useful shorthand for representing a number multiplied by itself many times. The phrase 2 multiplied by itself 20 times can be written as 2^{20}. The number *10* multiplied by itself 86 times can be written as 10^{86}. See appendix 3 for a review of exponents.

Information Is Closely Related to Probability

Suppose that the scientist has 100 coins, and he is told that he has a 1 in 64 chance of opening the door if he tosses the correct number of coins and enters an *H* or *T* into the computer. He is told not to toss all 100 coins because the combination for the door is not that long. He is also told that the door only has one correct combination.

How does the scientist figure out how many coins to toss? He needs to convert the odds of opening the door into an equivalent number of bits. One way is trial and error. He can compose a table like table 1.1. If the scientist tosses 6 coins, he will have a 1 in 64 chance of opening the door.

Table 1.1- Information Contained in Coins

Number of Coins	information in bits	possible outcomes	Odds
0	0	2^0=1	1 in 1
1	1	2^1=2	1 in 2
2	2	2^2=2x2 =4	1 in 4
3	3	2^3=2x2x2 =8	1 in 8
4	4	2^4=2x2x2x2 =16	1 in 16
5	5	2^5=2x2x2x2x2=32	1 in 32
6	**6**	**2^6= 2x2x2x2x2x2=64**	**1 in 64**
7	7	2^7 =2x2x2x2x2x2x2=128	1 in 128
8	8	2^8=2x2x2x2x2x2x2x2 =256	1 in 256
13	13	$2^{13=}$ 2x2x2x2.....x2= 8192	1 in 8192

Notice that there is a definite relationship between the number of bits in the door's combination and the odds that the scientist will open the door. Probability theory and information theory are closely related.

Mathematical Definition of Information

The following equation defines information:

Equation 1

$$2^{(\text{information})} = \frac{\text{Total Possible Outcomes}}{\text{Observed Outcome (s)}}$$

Example 1: a scientist is told to toss a coin 3 times and remember the results. How much information does he acquire when he observes the result? Answer: The total number of possible outcomes is 2x2x2 = 8, and only 1 outcome will be observed. So $2^{(\text{information})} = 8/1$. Because $2^3 = 2x2x2 = 8$, the scientist acquires 3 bits of information.

Example 2: Suppose the result in example 1 is H-H-T, but the scientist is unsure about the outcome because he cannot remember whether the first coin landed head or tails. How much information has he acquired? Answer: there are still 8 possible outcomes. The scientist observed either H-H-T or T-H-T, but he is not sure which. So both must be counted as observed outcomes. So $2^{(\text{information})} = 8/2 = 4$. Because $2^2 = 2x2 = 4$, the scientist acquires 2 bits of information.

Note that the odds of an event like winning the lottery are often expressed as one in some number, like a million. The one is the observed outcome, and the million represents the total number of possible outcomes. So the information acquired by knowing who wins this lottery is easy to calculate: $2^{(\text{information})}$ = (1 million outcomes / 1 outcome). Because $2^{20} = 1,048,576$, approximately 20 bits of information are acquired when the outcome of this lottery is observed.

The equation for information can be solved explicitly for information with the use of logarithms. Since many calculators have a log function, the next equation is often easier to use, but less intuitive than the first. See appendix three for a review of logarithms.

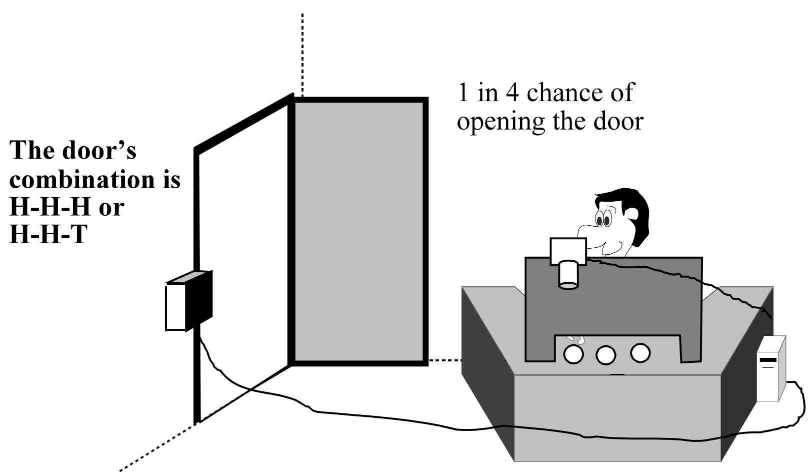
Equation 2

$$\text{information} = 3.32 \text{x} \log \left[\frac{\text{Total Possible Outcomes}}{\text{Observed Outcome (s)}} \right]$$

Trapped Scientist Who Cannot See the Results

Suppose that the scientist in figure 1.1 cannot see the results of the coin toss because a screen is placed between him and the coins (figure 1.2).

Figure 1.2: Trapped Scientist Who Cannot See the Results



In figure 1.2, the camera monitors the results of the coin toss and sends the results to the computer. Depending on the results of the coin toss**,** the computer is programed to do four things: open the door, close the door, beep once, and beep twice.

Coin 1	Coin 2	Coin 3	
H	H	H	--------->opens the door
H	H	T	--------->opens the door
H	T	H	--------->closes the door
H	T	T	--------->closes the door
T	H	H	--------->computer beeps twice
T	H	T	--------->computer beeps twice
T	T	H	--------->computer beeps once
T	T	T	--------->computer beeps once

The scientist cannot observe the results of the coins. He can only observe what the door and computer do after he tosses all three coins. The door opens for 2 of the 8 possible results. If the door is already open, and the camera observes H-H-H or H-H-T then the door will stay open. Two results will close the door if it is open, and have no effect if the door is already shut. Two results will cause the computer to beep once, and two results will cause the computer to beep twice. Does the scientist still acquire 3 bits of information when he tosses three coins?

Case 1: the door opens or stays open.
$2^{(\text{information})}$ = (8 possible outcomes/2 outcomes that cause this result) =4**.**
Since 2^2 =4, 2 bits of information are acquired.
Case 2: the door closes or stays closed.
$2^{(\text{information})}$ = (8 possible outcomes/2 outcomes that cause this result) =4**.**
So in the case, 2 bits of information are acquired.
Case 3: the computer beeps once, 2 bits of information are acquired.
Case 4: the computer beeps twice, 2 bits of information are acquired.

The average amount of information acquired each time the scientist tosses all 3 coins is now 2 bits. He is using 3 coins or 3 bits to transmit 2 bits of information. He must do this because the code that translates the result of the coin toss into what the door and computer do is not the optimal code. The optimal code should only require 2 coins to transmit 2 bits. One possible optimal code is as follows:

Coin 1	Coin 2	
H	H	--------->opens the door
H	T	--------->closes the door
T	H	--------->computer beeps once
T	T	--------->computer beeps twice

The average uncertainty per symbol (or coin in this example) is called the Shannon entropy. Shannon entropy* measures on average how much each observed symbol or coin decreases uncertainty. Because information corresponds to a reduction in uncertainty, Shannon entropy is also a measure of information. When 3 coins are used to transmit 2 bits (non-optimal code), the Shannon entropy is 2/3 of a bit per coin. With the optimal code, the Shannon entropy becomes 1 bit per coin. The total information transmitted in both cases is the same because 3 coins x 2/3 bit per coin = 2 coins x 1 bit per coin = 2 bits.

*Shannon entropy should not be confused with the term entropy as it is used in chemistry and physics. Shannon entropy does not depend on temperature. Therefore, it is not the same as thermodynamic entropy.[1] Shannon entropy is a more general term that can be used to reflect the uncertainty of any system. Thermodynamic entropy is confined to physical systems.

References:

1) Brillouin, Science and Information Theory, 1956.
2) Reza, An Introduction to Information, 1961.
3) Pierce, An Introduction to Information Theory, Symbols, Signals and Noise, 1961.
4) Yockey, Information Theory and Molecular Biology, 1992.