We have been seeking a mathematical model of a source of English text. Such a model should be capable of producing text which corresponds closely to actual English text, closely enough so that the problem of encoding and transmitting such text is essentially equivalent to the problem of encoding and transmitting actual English text. The mathematical properties of the model must be mathematically defined so that useful theorems can be proved concerning the encoding and transmission of the text is produces, theorems which are applicable to a high degree of approximation to the encoding of actual English text. It would, however, be asking too much to insist that the production of actual English text conform with mathematical exactitude to the operation of the model.

The mathematical model which Shannon adopted to represent the production of text (and of spoken and visual messages as well) is the *ergodic source.* To understand what an ergodic source is, we must first understand what a *stationary source* is, and to explain this is our next order of business.

The general idea of a stationary source is well conveyed by the name. Imagine, for instance, a process, i.e., an imaginary machine, that produces forever after it is started the sequences of characters

A E A E A E A E A E, etc.

Clearly, what comes later is like what has gone before, and *stationary* seems an apt designation of such a source of characters. We might contrast this with a source of characters which, after starting, produced

A E A A E E A A A E E E, etc.

Here the strings of A's and E's get longer and longer without end; certainly this is not a stationary source.

Similarly, a sequence of characters chosen at random with some assigned probabilities (the first-order letter approximation of example 1 above) constitutes a stationary source and so do the digram and trigram sources of examples 2 and 3. The general idea of a stationary source is clear enough. An adequate mathematical definition is a little more difficult.

The idea of stationarity of a source demands no change with time. Yet, consider a digram source, in which the probability of the second character depends on what the previous character is. If we start such a source out on the letter A, several different letters can follow, while if we start such a source out on the letter Q, the second letter must be U. In general, the manner of starting the source will influence the statistics of the sequence of characters produced, at least for some distance from the start.

To get around this, the mathematician says, let us not consider just one sequence of characters produced by the source. After all, our source is an imaginary machine, and we can quite well imagine that it has been started an infinite number of times, so as to produce an infinite number of sequences of characters. Such an infinite number of sequences is called an *ensemble* of sequences.

These sequences could be started in any specified manner. Thus, in the case of a digram source, we can if we wish start a fraction, 0.13, of the sequences with E (this is just the probability of E in English text), a fraction, 0.02, with W (the probability of W), and so on. *If we do this,* we will find that the fraction of E's is the same, averaging over all the *first* letters of the ensemble of sequences, as it is averaging over all the *second* letters of the ensemble, as it is averaging over all the *third* letters of the ensemble, and so on. No matter what position from the beginning we choose, the fraction of E's or of any other letter occurring in that position, taken over all the sequences in the ensemble, is the same. This independence with respect to position will be true also for the probability with which TH or WE occurs among the first, second, third, and subsequent *pairs* of letters in the sequences of the ensemble.

This is what we mean by stationarity. If we can find a way of assigning probabilities to the various starting conditions used in forming the ensemble of sequences of characters which we allow the source to produce, probabilities such that any statistic obtained by averaging over the ensemble doesn't depend on the distance from the start at which we take an average, then the source is said to be stationary. This may seem difficult or obscure to the reader, but the difficulty arises in giving a useful and exact mathematical form to an idea which would otherwise be mathematically useless.

In the argument above we have, in discussing the infinite ensemble of sequences produced by a source, considered averaging over-all *first* characters or over-all *second* or *third* characters (or pairs, or triples of characters, as other examples). Such an average is called an *ensemble* average. It is different from a sort of average we talked about earlier in this chapter, in which we lumped together all the characters in *one* sequence and took the average over them. Such an average is called a *time* average.

The time average and the ensemble average can be different. For instance, consider a source which starts a third of the time with A and produces alternately A and B, a third of the time with B and produces alternately B and A, and a third of the time with E and produces a string of E's. The possible sequences are

1. A B A B A B A B, etc.
2. B A B A B A B A, etc.
3. E E E E E E E E, etc.

We can see that this is a stationary source, yet we have the probabilities shown in Table V.

TABLE V

Probability of	*Time Average Sequence (1)*	*Time Average Sequence (2)*	*Time Average Sequence (3)*	*Ensemble Average*
A	½	½	0	⅓
B	½	½	0	⅓
E	0	0	1	⅓

When a source is stationary, and when every possible ensemble average (of letters, digrams, trigrams, etc.) is equal to the corresponding time average, the source is said to be ergodic. The theorems of information theory which are discussed in subsequent chapters apply to ergodic sources, and their proofs rest on the assumption that the message source is ergodic.[1]