They will be explored in some detail at the close of the chapter. But we must begin with information theory.

9.1 INFORMATION MEASURE: ENTROPY

The crux of information theory is the measure of information. Here we are using *information* as a technical term, not to be confused with its more conventional interpretations. In particular, the information of information theory has little to do with knowledge or meaning, concepts which defy precise definition, to say nothing of quantitative measurement. In the context of communication, information is simply that which is produced by the source for transfer to the user. This implies that before transmission, the information was not available at the destination; otherwise the transfer would be zero. Pursuing this line of reasoning, consider the following somewhat contrived situation.

A man is planning a trip to Chicago. To determine what clothes he should pack, he telephones the Chicago weather bureau and receives one of the following forecasts:

The sun will rise.
It will rain.
There will be a tornado.

Clearly, the amount of information gained from these messages is quite different. The first contains virtually no information, since we are reasonably sure in advance that the sun will rise; there is no uncertainty about this, and the call has been wasted. But the forecast of rain does provide information not previously available to the traveler, for rain is not an everyday occurrence. The third forecast contains even more information, tornadoes being relatively rare and unexpected events.

Note that the messages have been listed in order of decreasing likelihood and increasing information. The less likely the message, the more information it conveys to the user. We are thus inclined to say that information measure is related to *uncertainty*, the uncertainty of the user as to what the message will be. Moreover, the amount of information depends only on the message uncertainty, rather than its actual content or possible interpretations. Had the Chicago weather forecast been "The sun will rain tornadoes," it would convey information, being quite unlikely, but not much meaning.

Alternately, going to the transmitting end of a communication system, information measure is an indication of the *freedom of choice* exercised by the source in selecting a message. If the source can freely choose from many different messages, the user is highly uncertain as to which message will be selected. But if there is no choice at all, only one possible message, there is no uncertainty and hence no information.

Whether one prefers the uncertainty viewpoint or the freedom-of-choice interpretation, it is evident that the measure of information involves *probabilities*. Messages of high probability, indicating little uncertainty on the part of the user or little choice on the part of the source, convey a small amount of information, and vice versa. This notion is formalized by defining self-information in terms of probability.

Self-Information

Consider a source that produces various messages. Let one of the messages be designated A, and let P_A be the probability that A is selected for transmission. Consistent with our discussion above, we write the self-information associated with A as

$$\mathscr{I}_A = f(P_A)$$

where the function $f(\ )$ is to be determined. As a step toward finding $f(\ )$, intuitive reasoning suggests that the following requirements be imposed:

$$f(P_A) \geq 0 \qquad \text{where } 0 \leq P_A \leq 1 \tag{1}$$

$$\lim_{P_A \to 1} f(P_A) = 0 \tag{2}$$

$$f(P_A) > f(P_B) \qquad \text{for } P_A < P_B \tag{3}$$

The student should have little trouble interpreting these requirements.

Many functions satisfy Eqs. (1) to (3). The final and deciding factor comes from considering the transmission of *independent* messages. When message A is delivered, the user receives $\mathscr{I}_A$ units of information. If a second message B is also delivered, the total information received should be the sum of the self-informations, $\mathscr{I}_A + \mathscr{I}_B$. This summation rule is readily appreciated if we think of A and B as coming from different sources. But suppose both messages come from the same source; we can then speak of the compound message $C = AB$. If A and B are statistically independent, $P_C = P_A P_B$ and $\mathscr{I}_C = f(P_A P_B)$. But the received information is still $\mathscr{I}_C = \mathscr{I}_A + \mathscr{I}_B = f(P_A) + f(P_B)$ and therefore

$$f(P_A P_B) = f(P_A) + f(P_B) \tag{4}$$

which is our final requirement for $f(\ )$.

There is one and only one function† satisfying the conditions (1) to (4), namely, the *logarithmic function* $f(\ ) = -\log_b(\ )$, where b is the logarithmic base. Thus self-information is defined as

$$\mathscr{I}_A \triangleq -\log_b P_A = \log_b \frac{1}{P_A} \tag{5}$$

† See Ash (1965, chap. 1) for proof.