CHAPTER 5

INFORMATION OF EVENTS WITH DISCRETE OUTCOMES: APPLICATIONS IN COMMUNICATIONS SCIENCE AND IN PSYCHOLOGY

COMMUNICATIONS SCIENCE

We have now developed some of the fundamental equations governing information transmitted in the presence of noise. The type of source we have considered is one in which successive symbols emitted are statistically independent of each other. For example, if the alphabet of transmitted symbols is just the ordinary English alphabet, the probability $p(h)$ of emitting the letter h from the source is independent of the previous letter emitted. However, experience tells us that this assumption of independence is unrealistic. For example if the preceding letter was s or c, the probability of h following would seem to be greater than if the preceding letter was, say, a or b, since there are, surely, more words containing the combination *sh* or *ch* than *ah* or *bh*. The type of source we have modeled is called a *zero memory source*. Suppose that the 26 letters of the English alphabet and the "blank" symbol are equally probable. An upper limit to the source entropy by Equation (4.4) is, then,

$$H(X) = \log_2 27 = 4.76 \text{ bits per letter.}$$

In fact, a more realistic estimate places $H(X)$ between 0.6 and 1.3 bits per letter (Pierce, 1980). Nonetheless, bearing in mind the limitations of the zero memory model, we can begin to appreciate the applicability of our equations.

The application of information theory by communications engineers is based on the equivalence of the *bit* with the *binary digit*. If a noiseless channel can transmit C on-or-off pulses per second, then it can transmit C binary digits per second; and if each binary digit carries one bit of information, the channel can transmit C bits per second. We say that the *channel capacity* is C bits per second.

Suppose the source entropy for the noiseless channel is $H(X)$ bits per symbol. Then if the channel capacity is C bits per second, the output of the source can be *coded* in such a way that the channel can transmit an average of C/H symbols per second. The average rate of transmission cannot exceed this value of C/H. The above theorem is due to Shannon.

We cannot discuss here the science and art of coding; it would take us too far afield. The reader is referred to the texts on information theory, which treat the matter of coding very thoroughly.

For the case of the discrete noisy channel we refer to Equation (4.21),

$$\mathcal{I}(X|Y) = H(X) - H(X|Y) \,. \tag{4.21}$$

Suppose that all quantities in this equation are measured in units of bits per second. The channel capacity, C, is equal to the maximum possible value of the rate of transmission for the channel, $\mathcal{I}(X|Y)_{\max}$ [bits per second]. Shannon's fundamental theorem for a noisy channel is as follows. Choose a value of $\mathcal{I}(X|Y) = \mathcal{I}$. If $\mathcal{I} < C$ there exists a coding system such that the output of the source can be transmitted over the channel with an arbitrarily small frequency of errors (or an arbitrarily small equivocation).

Clearly, the above is of great interest to communications engineers. A primary application of information theory in communications science is, therefore, the calculation of channel capacity. Knowledge of channel capacity, then, leads to the development of efficient codes for the transmission of messages.

Perhaps the real power of information theory in communications comes with the analysis of channels that transmit continuous rather than discrete signals; that is, continuous wave forms rather than just discrete zeros and ones and the like. For such continuous or "analog" channels, using the known ratio of signal-to-noise power and the bandwidth of the channel, one can calculate *a priori* the value of the channel capacity. Therefore, one can calculate the greatest number of bits per second transmissible by the channel from a knowledge of two physical parameters. The nature of such a calculation awaits our introduction to the information of events with continuous outcomes (Chapter 8), but we can certainly appreciate its significance even at this stage.

PSYCHOLOGY: CATEGORICAL JUDGMENTS

We have seen in Chapter 4 the manner in which psychologists utilized the information transmission function [Equations (4.21) and (4.33)] to calculate the amount of information (in bits per stimulus) transmitted in an experiment on categorical judgments. Many different types of categorical experiment were performed evaluating $\mathcal{I}(X|Y)$, for example, for brightness, hue, position of a pointer in a linear interval, loudness, pitch, odor intensity, taste intensity, etc. Stimuli of the simplest type, such as loudness or pitch, have been termed "one-dimensional," while position of a dot in a square would be called "two-dimensional" because both horizontal and vertical coordinates must be identified by the subject. Similarly, identification of colors of equal luminance has been termed two-dimensional, since the colors can vary both in hue and in saturation. We shall confine our discussion to one-dimensional stimuli.

In Chapter 4, we discussed the general theory of measuring transmitted information, $\mathcal{I}(X|Y)$, for an experiment on categorical recognition or judgment. Let us now take a specific numerical example of a stimulus-response or confusion matrix, and evaluate $\mathcal{I}(X|Y)$ using Equation (4.21) or (4.21a). Consider the hypothetical experiment whose stimulus-response matrix is given in Table 5.1. Dividing each matrix element by the sum of all the elements, which is 125, will give the matrix of joint probabilities, Table 5.2. Table 5.3 gives the matrix of conditional probabilities $p(y_k|x_j)$, calculated from Equation (4.30). Also entered in this table are $p(x_j)$ and $p(y_k)$, evaluated from Equations (4.27) and (4.28). Table 5.4 gives the matrix of conditional probabilities $p(x_j|y_k)$, calculated from Equation (4.29). From Equation (4.19),

$$H(X) = -5(0.2\ln 0.2) = 1.6094 \text{ natural units.}$$

Table 5.1 Stimulus-Response Matrix

Stimulus categories	Response Categories y_1	y_2	y_3	y_4	y_5	$N_{j\cdot}$
x_1	20	5	0	0	0	25
x_2	5	15	5	0	0	25
x_3	0	6	17	2	0	25
x_4	0	0	5	12	8	25
x_5	0	0	0	6	19	25
$N_{\cdot k}$	25	26	27	20	27	125

Table 5.2 Matrix of Joint Probabilities $p(x_j, y_k)$

	y_1	y_2	y_3	y_4	y_5	$p(x_j)$
x_1	0.16	0.04	0	0	0	0.20
x_2	0.04	0.12	0.04	0	0	0.20
x_3	0	0.048	0.136	0.016	0	0.20
x_4	0	0	0.04	0.096	0.064	0.20
x_5	0	0	0	0.048	0.152	0.20
$p(y_k)$	0.20	0.208	0.216	0.16	0.216	1.00

Table 5.3 Matrix of Conditional Probabilities $p(y_k|x_j)$

	y_1	y_2	y_3	y_4	y_5	$p(x_j) = N_{j.}/125$
x_1	0.80	0.20	0	0	0	0.20
x_2	0.20	0.60	0.20	0	0	0.20
x_3	0	0.24	0.68	0.08	0	0.20
x_4	0	0	0.20	0.48	0.32	0.20
x_5	0	0	0	0.24	0.76	0.20
$p(y_k) = N_{.k}/125$	0.200	0.208	0.216	0.160	0.216	1.00

Table 5.4 Matrix of Conditional Probabilities $p(x_j|y_k)$

	y_1	y_2	y_3	y_4	y_5
x_1	0.80	0.19231	0	0	0
x_2	0.20	0.57692	0.185185	0	0
x_3	0	0.23077	0.629629	0.1	0
x_4	0	0	0.185185	0.6	0.296296
x_5	0	0	0	0.3	0.703703

Multiplying the elements of Table 5.2 by the logarithms of the corresponding elements of Table 5.4 (recalling that $0 \log 0 = 0$) gives $H(X|Y)$ [Equation (4.20)].

$$H(X|Y) = 0.77518 \text{ natural units.}$$

From Equation (4.21), transmitted information per stimulus

$$\mathscr{I}(X|Y) = 1.6094 - 0.77518 = 0.83426 \text{ natural units.}$$
$$= 0.83426 / \ln 2 = 1.2036 \text{ bits.}$$

The above represents rather a lot of work, even with the use of a calculator, although there were only 5 categories. Alternatively, a spreadsheet can be used with calculations performed using macros. Once you understand the principles involved, you may wish to use a computer program like the one written in Basic provided below. The program uses standard matrix notation, so that $C(I, J)$ represents N_{ij}, which is the number in the i^{th} row and j^{th} column of the original stimulus-response matrix.

```
10 ' MARCH 26, 1991; K. H. NORWICH
20 ' CALCULATES TRANSMITTED INFORMATION FROM A CONFUSION MATRIX.
30 ' INFORMATION, IXGY =IYGX, OR I(X|Y)=I(Y|X), IS CALCULATED FROM
40 ' EQUATION 4.21.
50 '
60 INPUT "NUMBER OF ROWS = ";M
70 INPUT "NUMBER OF COLUMNS = ";N
80 DIM C(M,N):' CONFUSION MATRIX
90 DIM PJ(M,N):' JOINT PROBABILITY MATRIX
100 DIM PXGY(M,N):' MATRIX OF CONDITIONAL PROBABILITIES, X Given Y
110 DIM PYGX(M,N):' MATRIX OF CONDITIONAL PROBABILITIES, Y Given X
120 DIM SUMX(N):' N.k
130 DIM SUMY(M):' Nj.
140 '
150 PRINT "ENTER DATA FOR PRIMARY CONFUSION MATRIX"
160 FOR X=1 TO M
170    FOR Y=1 TO N
180       PRINT "C(";X;Y;") = ";
190       INPUT C(X,Y)
200       TOTAL = TOTAL + C(X,Y)
```

```
210    NEXT Y
220 NEXT X
230 '
240 'CALCULATE Nj. AND N.k AND JOINT PROBABILITY MATRIX
250 FOR X = 1 TO M
260    FOR Y = 1 TO N
270       SUMY(X) = SUMY(X) + C(X,Y)
280       SUMX(Y) = SUMX(Y) + C(X,Y)
290       PJ(X,Y) = C(X,Y)/TOTAL
300     NEXT Y
310 NEXT X
320 '
330 ' CALCULATE SOURCE AND RECEIVER ENTROPIES
340 FOR X = 1 TO M
350     HX = HX - (SUMY(X)/TOTAL) * LOG(SUMY(X)/TOTAL)
360 NEXT X
370 FOR Y = 1 TO N
380     HY = HY - (SUMX(Y)/TOTAL) * LOG(SUMX(Y)/TOTAL)
390 NEXT Y
400 '
410 ' CALCULATE SOURCE AND RECEIVER ENTROPIES AND MATRICES OF CONDITIONAL PROBS
420 FOR X = 1 TO M
430     FOR Y = 1 TO N
440         PYGX(X,Y) = C(X,Y)/SUMY(X)
450         PXGY(X,Y) = C(X,Y)/SUMX(Y)
460     NEXT Y
470 NEXT X
480 '
490 ' CALCULATE H(X|Y) AND H(Y|X)
500 FOR X = 1 TO M
510     FOR Y = 1 TO N
520         IF PXGY(X,Y) = 0 THEN 540
530         HXGY = HXGY - PJ(X,Y) * LOG(PXGY(X,Y))
540         IF PYGX(X,Y) =0 THEN 560
550         HYGX = HYGX - PJ(X,Y) * LOG(PYGX(X,Y))
560     NEXT Y
570 NEXT X
580 '
590 ' CALCULATE INFORMATION BY TWO FORMULAS
600 IXGY = HX - HXGY: ' EQUATION(4.21)
610 IYGX = HY - HYGX: ' EQUATION(4.36)
620 '
630 ' PRINT OUT SOURCE AND RECEIVER ENTROPIES,
640 ' EQUIVOCATIONS AND INFORMATION TRANSMITTED, I(X|Y) = I(Y|X)
650 PRINT "SOURCE ENTROPY = "; HX; " NATURAL UNITS, OR "; HX/LOG(2); " BITS"
660 PRINT "RECEIVER ENTROPY = "; HY; "NATURAL UNITS, OR "; HY/LOG(2); " BITS"
670 PRINT "SOURCE EQUIVOCATION = "; HXGY; "NATURAL UNITS, OR"; HXGY/LOG(2); "
    BITS"
680 PRINT "RECEIVER EQUIVOCATION = "; HYGX; "NATURAL UNITS, OR"; HYGX/LOG(2); "
    BITS"
690 PRINT "I(X|Y) = "; IXGY; " NATURAL UNITS, OR" ;IXGY/LOG(2); " BITS"
700 PRINT "I(Y|X) = "; IYGX; " NATURAL UNITS, OR" ;IYGX/LOG(2); " BITS"
710 END
```

Both for fun and further insight, you might like to consider the following problem. Suppose that in an experiment on categorical judgments there are n categories. In the resulting stimulus-response matrix there are only n non-zero elements. No two of the non-zero elements occupy the same row or the same column. Suppose, further, that all of the diagonal elements, N_{ii}, are zero. That is, our hapless subject did not correctly identify even one stimulus! For example, if $n = 3$, the matrix may look like this:

$$\begin{array}{r|ccc} & \multicolumn{3}{c}{\textit{Response } y} \\ \hline & 0 & a & 0 \\ \textit{Stimulus } x & 0 & 0 & b \\ & c & 0 & 0 \end{array} \qquad a, b, c > 0$$

Every time a stimulus from category 1 was given, it was consistently identified as belonging to category 2; every time a stimulus from category 2 was given, it was consistently identified as belonging to category 3, etc.

What values would you expect for the equivocations $H(X|Y)$ and $H(y|X)$? Derive the equivocations mathematically or calculate them using the computer program. What does this strange result mean?

Turning now to another matter, how will $\mathcal{I}(X|Y)$, the measured transmitted information, vary with the number of categories, n, in an experiment on categorical judgments of one-dimensional stimuli? Let us regard the stimulus probabilities, $p(x_j)$, since they are under control of the experimenter, as equally probable so that $p(x_j) = 1/n$. Therefore, the source or stimulus entropy, $H(X)$ (Equation (4.19)), is equal to $\log_2 n$ bits. Let us also select the range of stimuli to extend from the minimum perceptible stimulus to the maximum perceptible stimulus (below the threshold for pain). We can now draw a graph of transmitted information, $\mathcal{I}(X|Y)$, against stimulus entropy, $H(X)$. We should recall here Equation (4.21):

$$\mathcal{I}(X|Y) = H(X) - H(X|Y)\,. \tag{4.21}$$

Clearly, when $n = 1$, $H(X) = 0$ and $\mathcal{I}(X|Y) = 0$; if there is only one category, no information can be transmitted since there was no prior uncertainty. Suppose that $n = 2$, and that the two stimulus categories lie at opposite ends of the spectrum of stimulus values, category 1 bordering the threshold value, and category 2 bordering the greatest possible stimulus value if this can be defined. For stimuli of the intensity type, the greatest stimulus intensity would be close to the maximum allowable physiological value. In this experiment ($n = 2$), it is unlikely that stimuli would be confused. The equivocation would be close to zero, so that $\mathcal{I}(X|Y) = H(X) \cong 1$ bit. Thus, for the first two values of n considered, $\mathcal{I}(X|Y)$ has been equal or very nearly equal to $H(X)$.

Now let us increase n to 3 by adding a category between the high and low extremes. Will the subject make any mistakes in identifying categories? Occasionally, perhaps. However, with n equal to 4 categories, mistakes will probably be made; $H(X|Y)$ will be greater than zero and $\mathcal{I}(X|Y)$ will be less than $H(X)$. As n increases, $\mathcal{I}(X|Y)$ will continue to rise, but less than the rise in $H(X)$ and, thus, will fall below the straight line $\mathcal{I}(X|Y) = H(X)$. The graph obtained experimentally is shown schematically in Figure 5.1. $\mathcal{I}(X|Y)$ approaches an asymptote at about 2.5 bits of information per stimulus ($\mathcal{I}(X|Y)_{\max}$).

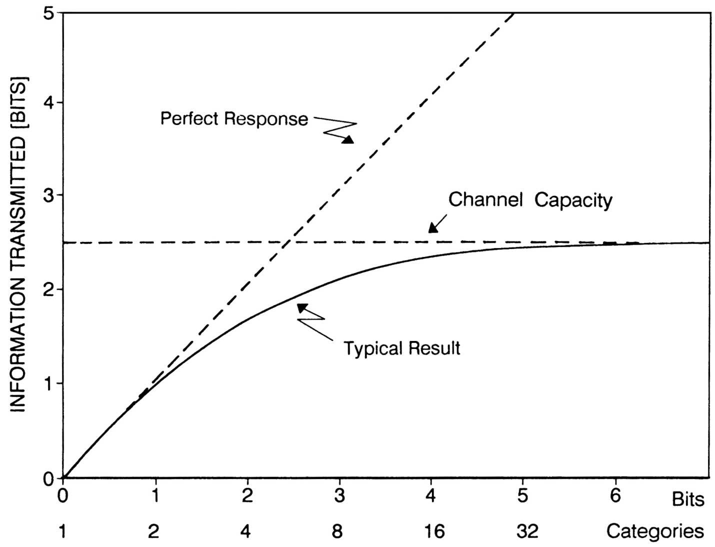

Figure 5.1 Schematic graph representing the information transmitted in an experiment on categorical judgments. The abscissa contains the stimulus entropy, $H(X)$, measured in number of equally probable categories, and in bits as $\log_2$ (*number of equally probable categories*). The ordinate contains the information transmitted, $\mathcal{I}(X|Y)$, as calculated from Equation (4.21). The straight line $\mathcal{I}(X|Y) = H(X)$ represents the "best" possible performance, with equivocation equal to zero. Typically, the experimentally obtained curve falls progressively below the line of best performance and approaches $\mathcal{I}(X|Y)_{\max}$, which is the channel capacity, drawn at about 2.5 bits per stimulus.

While the position of this asymptote is not invariant, it does not change dramatically for different types of stimulus. For example, for pitch $\mathscr{I}(X|Y)_{\max}$ is about 2.5 bits per stimulus (Pollack, 1952, 1953); for auditory intensity, 2.3 - 2.5 bits per stimulus (Garner, 1953); and for taste 1.7 - 1.9 bits per stimulus (Beebe-Center *et al.*, 1955). When you think about it, this is quite surprising. Why should there be a common upper limit to our ability to make "absolute" or categorical judgments? If we set $H(X|Y)$ equal to zero in Equation (4.21) we can write

$$\mathscr{I}(X|Y)_{\max} = H(X)_{\max} - 0 \,. \tag{5.1}$$

Suppose we also set

$$H(X)_{\max} = \log_2(n_{\max}) \,, \tag{5.2}$$

since stimulus categories are equally probable. Therefore

$$n_{\max} = 2^{\mathscr{I}(X|Y)\max} = 2^{2.5} = 5.7 \text{ categories.} \tag{5.3}$$

That is, it is "as if" 5.7 (equally probable) categories can be identified without error. Hence Miller's famous paper "The magical number seven, plus or minus two: some limits on our capacity for processing information" (1956). Later studies, particularly on categorical judgments of "intensities," might suggest that the number is closer to 6 than 7, and, in a flight of fancy, we shall later speculate on the merits of $2\pi \simeq 6.3$.

The value of $\mathscr{I}(X|Y)$ has been termed the *channel capacity* by analogy with the use of this term in communications science. The measurement of channel capacity for categorical judgments has been subjected to a great deal of analysis. For example, "edge effects" must be considered: the enhanced ability of subjects to categorize stimuli at the extremes values. "Practice effect" refers to the ability of subjects to increase $\mathscr{I}(X|Y)_{\max}$ by prolonged practice or to the augmented channel capacity of experts; for example, it seems that wine tasters possess an enhanced ability to identify tastes. "Feedback effect" refers to the subject's being told, after making a judgment, whether it was correct, etc. Anomolous results may emerge if $\mathscr{I}(X|Y)_{\max}$ is computed from an averaged confusion matrix rather that a matrix representing a subject's performance on a single occasion.

Values of $\mathscr{I}(X|Y)_{\max}$ measured from two-dimensional stimuli are greater than those measured from one-dimensional stimuli, but usually not twice as great. For example, $\mathscr{I}(X|Y)_{\max}$ for a pointer in a linear interval has been measured to be 3.25 bits (Hake and Garner, 1951), while $\mathscr{I}(X|Y)_{\max}$ for a pointer in a square equals about 4.6 bits $\neq 2 \times 3.25$ bits (Klemmer and Frick, 1953).

I shall have a good deal to say in this book about one-dimensional stimuli of the intensity type, such as the intensity of sound or the concentration of a solution that is to be tasted. I shall approach the question of the maximum information transmitted per stimulus in quite a different way; that is, I shall suggest a rather different way of measuring the channel capacity.

THE CONCEPTUAL BASIS FOR THE APPLICATION OF INFORMATION THEORY AS DEVELOPED IN THE 1950'S

To review:

In the 1950's, both communications engineers and psychologists began to utilize information theory to calculate channel capacities. Stripped to its conceptual bones, the process was as follows. An entity called "information," whose magnitude could be calculated from the properties of a channel of communication, was defined. Each channel was tagged or labeled with a number called the "channel capacity," which represented the greatest amount of information that could be transmitted per symbol (or per stimulus) using this channel. This channel capacity was then utilized in different ways by the respective disciplines, the engineers used it as a guide for coding, the psychologists as a means of analyzing perceptual processes.

Two features of information theory as it was employed by these disciplines should be recognized: (i) its non-unique nature and (ii) its extrinsic nature. These two properties, as we shall see, are not totally independent.

(i) Shannon's seminal paper on what is now called "information theory" was entitled, modestly, "*A* mathematical theory of communication"; it was not entitled "*The* mathematical theory of communication." There was wisdom in Shannon's selection of the indefinite article, since there are other possible measures of information. Perhaps the most celebrated is *Fisher's information* (Fisher, 1949). Fisher defined information to be proportional to the reciprocal of variance. For example, a calculated quantity is subject to random error and, as a result, has variance σ^2. If the experiment or measurement is then repeated N times, the variance will be reduced to σ^2/N. Therefore, for large N, variance becomes small. The reciprocal of variance is, therefore, a measure of information provided by measurements (see, for example, Bell, 1968). There are other, more general measures of entropy and information (see, for example, Renyi, 1970; van der Lubbe *et al*., 1984). Many people, myself included, think that Shannon's measure is the most natural and most useful but, admittedly, it is not unique. Consider also *algorithmic information content* or *algorithmic randomness* of a physical entity, which is defined by Zurek (1990) as "the size (in bits) of the most concise message (e.g. of the shortest program for a universal computer) which describes that entity with requisite accuracy."

(ii) By "extrinsic" I refer to the tag or label that is pinned on the communications channel, branding it with a stated channel capacity. Because there are other possible measures of information it is, in principle, possible to label the channel with a different number, representing an alternative measure. The point is, though, that our calculations or predictions of the physical operation of the channel are unaffected by the information label we put upon it. That is, the physical theory of channel function is independent of which mathematical equation we use to assess its efficiency in transmitting information. The physical system is, so to speak, oblivious of the measure of information used by the engineer or psychologist to appraise it.

There is a tendency across all disciplines by those who apply the information measure to do so in a manner in which the measure is non-unique and extrinsic. For example, Norbert Wiener, in his famous book *Cybernetics* (1948), introduced the world to the logarithmic measure of information with the following sentences: "We know *a priori* that a variable lies between 0 and 1, and *a posteriori* that it lies on the interval (a, b) inside (0, 1). The the amount of information we have from our *a posteriori* knowledge is

$$-\log_2 \frac{\text{measure of }(a, b)}{\text{measure of }(0, 1)}\ ."$$

The measure is non-unique: we could select other measures of the amount of information given by knowing the position of a point in a line. The measure is also extrinsic to the line interval (a, b), in the sense that we could adopt a different measure of information, and this would not in any way affect our measure of the interval (a, b). The measure of (a, b) is totally unaffected by the method we use to assess its information content. The non-uniqueness and extrinsic properties are related.

G. Gatlin (1972) advocated a measure of the information content of a chain of DNA. The maximum possible entropy of a length of DNA is calculated *a priori* by assuming that at each position in the molecule the four bases adenine, guanine, thymine and cytosine were equally probable. The *a posteriori* entropy is then calculated after analyzing the DNA molecule and measuring the actual base frequencies. The difference between *a priori* value and *a posteriori* value is then taken as a measure of the information content of the molecule. Again the measure is non-unique, since we could devise other measures of information content of DNA (e.g. the number of amino acids in the corresponding protein structure?), and extrinsic because the biochemistry of DNA does not seem to depend on or change with an information measure.

All this is, of course, not surprising, but I raise the issue here in order to contrast the information scientist's use of the information measure with the physicist's use of the information measure. We shall see in the next chapter that the physicist's measure of information is, indeed, (i) unique and (ii) intrinsic to the system. That is, the physicist may not make an arbitrary selection from among the various mathematical functions that measure information, and, if he or she should select an alternative measure, the calculated or predicted behavior of the system under study (cf. communications channel) would differ from the observed behavior. The calculated information enters as a variable into the dynamics of the physical system.

REFERENCES

Beebe-Center, J.G., Rogers, M.S. and O'Connell, D.N. 1955. Transmission of information about sucrose and saline solutions through the sense of taste. *Journal of Psychology*, **39**, 157-160.

Bell, D.A. 1968. *Information Theory and it Engineering Applications*, 4th edition. Pitman and Sons, London.

Fisher, R.A. 1949. *The Design of Experiments*, 5th edition, page 182. Oliver and Boyd.

Garner, W.R. 1953. An informational analysis of absolute judgements of loudness. *Journal of Experimental Psychology* **46**, 373-380.

Gatlin, G. 1972. *Information Theory and Living Systems*. Columbia University Press, New York.

Hake, H.W. and Garner, W.R. 1951. The effect of presenting various numbers of discrete steps on scale reading accuracy. *Journal of Experimental Psychology* **42**, 538-566.

Klemmer, E.T. and Frick, F.C. 1953. Assimilation of information from dot and matrix patterns. *Journal of Experimental Psychology* **45**, 15-19.

Miller, G.A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* **63**, 81-97.

Pierce, J.R. 1980. *An Introduction to Information Theory: Symbols, Signals and Noise*. page 103, 2nd edition, Dover, New York.

Pollack, I. 1952. The information of elementary auditory displays. *Journal of the Acoustical Society of America*, **24**, 745-749.

Pollack, I. 1953. The information of elementary auditory displays II. *Journal of the Acoustical Society of America*, **25**, 765-769.

Renyi, A. 1970. *Probability Theory*. IX, §6, North-Holland, Amsterdam.

Van der Lubbe, J.C.A., Boxma, Y. and Boekee, D.E. 1984. A generalized class of certainty and information measures. *Information Sciences* **32**, 187-215.

Wiener, N. 1948. *Cybernetics*. Wiley, New York.

Zurek, W.H. 1990. Algorithmic information content, Church-Turing thesis, physical entropy and Maxwell's demon. In: *Complexity, Entropy and the Physics of Information*, W.H. Zurek, Ed. p 73-89. Addison-Wesley, Redwood City, California.