Information, sensation, and perception by K H Norwich; Academic Press, San Diego, CA, 1993, 326 pages, $59.95 (£46.00) ISBN 0 12 521890 7

One might expect a book under this title to be simply an application of information theory, once very fashionable in psychology, to problems of sensation and perception. But the author declares that his ultimate concern is with the epistemological intuitions of George Berkeley, the 18th century English philosopher; information theory enters merely as a vehicle for expressing those intuitions in operational terms. Present-day perceptual psychologists, however, will be interested chiefly in the author's attempt to accommodate some ten empirical relationships ('laws') within a coherent mathematical theory, a theory consciously constructed after the pattern of physical science, borrowing especially from the ideal communication theory of Shannon (1948) and also from thermodynamics.

Mathematical argument is more precise than verbal, permitting thereby the formulation of much more delicate arguments, but it is otherwise no different in kind. At the end of the day it has to make psychological sense, such as any verbal argument. The author may choose his symbols to signify what he pleases, but thereafter each formula is a precise expression which must transpose into some meaningful psychological statement. However, many readers will be sufficiently unsure in following the algebraic manipulations to be confident that their interpretation of the formulae is accurate. The chief task of this review, therefore, is to decode the principal equations and express their meanings in psychological terms. This is not a trivial matter.

At the heart of the mathematical argument is the equation (page 140)

\[ F = \frac{1}{2} k \ln(1 + \beta I^n/t). \]  

(1)

Here \( F \) is the strength of sensation, \( k \) a constant of proportionality, \( \beta \) another constant, \( I \) the stimulus intensity in physical units, \( n \) an exponent, and \( t \) the duration of the stimulus. If \( \beta I^n/t > 1 \), then

\[ F = \frac{1}{2} k [\ln(\beta/t) + n \ln I], \]  

(2)

which is Fechner's Law. If \( \beta I^n/t < 1 \), then

\[ F = \frac{1}{2} k \beta I^n/t, \]  

(3)

which is Stevens's Power Law. This is presented (page 151) as a unification of the two principal proposals for a psychophysical law; but that appearance is deceptive, as the author himself seems to realise on pages 253–254.

Equation (1) is a compound of two unrelated ideas; first,

\[ F = \frac{1}{2} k \ln(1 + S/N), \]  

(4)

which is Shannon's (1948) formula for the information transmitted by a signal of power \( S \) through a noisy channel (noise level \( N \)) and, second,

\[ S = \beta N I^n/t. \]  

(5)

Stevens's Power Law [equation (3)] derives directly from equation (5) and is assumed, not explained. Moreover, \( I^n \) is said to be "the variance of the stimulus population", so that, if \( I \) measures luminance (a Poisson process), \( n \) should, presumably, be 1.

The independence of the present treatments of Fechner's and Stevens's Laws is brought out in a comparison with Teghtsoonian's (1971) proposed integration. Teghtsoonian proposed that the Power Law mapped each physical continuum onto a common internal scale of sensation on which all discriminations and so on were realised. All Weber fractions should therefore be the same when expressed in subjective units, and should differ at the physical level of description only by virtue of their different power law exponents. This leads to the relation

\[ (1 + A S/S)^n = \text{constant} \]  

(6)

for any stimulus continuum \( S \) with exponent \( n \), a relation between the Weber fraction and the
Power Law exponent which (if Teghtsoonian is correct) should apply to all stimulus continua conforming to Weber's Law. The truth of equation (6) can be subjected to empirical examination. In the present theory there is no corresponding relationship which might be tested.

The author does, however, use equation (5) to relate Stevens's Power Law to Pieron's Law for simple reaction time by saying that the exponents in the two laws are the same. But the value of the exponent for, say, loudness varies so much from one experiment to another [Marks (1974) has assembled estimates, calculated with respect to amplitude, ranging from 0.37 to 0.85] that, to my mind, approximate numerical equality is far from sufficient. A more detailed justification is required.

For the remainder of this review I shall confine attention to the author's use of equation (4), which is the more substantial part of his thesis. To prepare the reader, there are four chapters on information theory, treating both discrete and continuous probability distributions. Those chapters rely overmuch on pages of algebraic formulae to express ideas and generally lack the prior verbal explanations which would have made things easier for the reader. In particular, there is no characterisation at all of the fundamental notion of statistical information. I will endeavour to supply, in the brief space available, the explanation that the book omits.

Suppose I conduct an experiment and collect some data. I then carry out a statistical test to see whether my data accord with some prior hypothesis or not. The Neyman-Pearson lemma states that the optimum statistical test depends on a bisection of the probability ratio,

\[ P(Data|H_1)/P(Data|H_0), \]

and Kullback (1959) has shown that all the usual parametric tests can be formulated as bisections of likelihood-ratio with respect to suitably chosen null \( (H_0) \) and alternative \( (H_1) \) hypotheses.

Suppose I now repeat my experiment, seeking a more sensitive result. Informally, I add the information from the two replications together. Since

\[ P(Data_1 \cap Data_2|H_0) = P(Data_1|H_0) \times P(Data_2|H_0), \]

(independent probabilities multiply),

\[
\ln\left(\frac{P(Data_1 \cap Data_2|H_1)}{P(Data_1 \cap Data_2|H_0)}\right) = \ln\left(\frac{P(Data_1|H_1)P(Data_2|H_1)}{P(Data_1|H_0)P(Data_2|H_0)}\right) + \ln\left(\frac{P(Data_2|H_1)}{P(Data_2|H_0)}\right) + \ln\left(\frac{P(Data_1|H_1)}{P(Data_1|H_0)}\right)
\]

(independent log probability ratios add). It is therefore natural to take log probability ratio as the measure of the information afforded by the experiment (as Kullback 1959).

Note that the fundamental measure has to be a log probability ratio. If only one probability (one hypothesis, \( H_0 \)) is considered, one gets an expression of uncertainty (analogous to the entropy of a thermodynamic system) which, for a continuous distribution, is not independent of the underlying metric. That dependence disappears in a probability ratio. Uncertainty is analogous to the velocity potential in fluid motion or voltage in an electrical circuit; only differences in potential or voltage are physically significant. This problem surfaces in chapters 7 and 8 and the author's treatment stumbles.

This brings me to the point of my digression, that the appropriate measure of information depends on two hypothetical states of nature, the two which are implicitly compared in the statistical test. A simple-minded argument would say that (a) human subjects process information, (b) formula (4) measures information, and (c) must therefore be fundamental to the underlying process. But information cannot be quantified absolutely; it is 'about something'—it has to be measured in relation to what it is about. What then are the two alternative states of nature implicit in Shannon's measure of information on which the argument depends?

One interpretation of Shannon's formula runs as follows: suppose a voltage \( x \) is transmitted through a communication channel and received as a voltage \( y \). The channel might be open circuit (null hypothesis), in which case \( x \) and \( y \) will be independent normal variables, or it might simply be subject to interference by noise. The specific alternative considered by Shannon is that voltage \( y \) is equal to voltage \( x \) plus an independent noise voltage. If that noise voltage has variance \( N \) and the input voltage variance \( S \), then the information transmitted is \( \frac{1}{2}\ln(1+S/N) \) (Kullback 1959, page 9); except that if the channel has bandwidth \( W \), \( 2W \) independent voltages can be transmitted per second (Shannon 1949) and the information throughput is \( W\ln(1+S/N) \).
In transmitting messages through a real communication channel, it is plausible that the amount of information transmitted (about the fidelity of the channel) should increase in direct proportion to the length of the message and thereby provide a theoretical measure of (minimum) message length. This explains the relevance of information theory to ideal communication systems, but that is not the present context.

In the present context formula (4) needs to be interpreted as the information in favour of a stimulus of variance $S$ (strictly $S + N$) against the alternative of no stimulus at all (variance $N$). If the stimulus be increased from $S$ to $S + \Delta S$, then

$$\Delta F = \frac{1}{2} k \ln \left( 1 + \frac{\Delta S}{S + N} \right)$$

is the information in favour of $S + \Delta S$ against $S$, and this is precisely the kind of discrimination which subjects are asked to make. The difference is detected if $\Delta F$ exceeds a minimum value; this condition is equivalent to

$$\frac{\Delta S}{S + N} = \text{constant}$$

which is Weber's Law. Formula (10) [and also (4)] is of the form $\ln(\Delta S + S + N) - \ln(\Delta S + N)$, which is symmetric in the two stimulus magnitudes, $S$ and $S + \Delta S$; this means that it measures the information for discrimination between two separate stimuli presented for the same period of time—mode 1' (page 207)—which also happens to be the stimulus configuration for which Weber’s law holds well. But the author believes that his mathematics applies to a different mode in which a brief increment is superimposed on a continuous background stimulus, and his example data are of that kind. The psychophysical properties of these two modes are systematically different, a matter of which the author is aware.

One further point is important. Formula (4) is derived on the assumption that $S$ and $N$ are the variances of Gaussian variables of zero mean (for example, signal voltages in a communication line). But a visual stimulus, say a flash of light of luminance $L$, is equivalent at the physical level of description to a Poisson process. For not too low a luminance the variation of a Poisson variable about its mean may be modelled as a Gaussian (of variance $L$). But the Poisson variable also has a mean, again equal to $L$, and changes in the mean are much more informative than changes in the variance. What happens to the information supplied by the mean luminance? It does not figure anywhere in the book.

There is an implicit, unwitting, assumption that the sensory system is differentially coupled to the physical stimulus so that the mean input is not transmitted. Only changes in the stimulus input, including those occurring by chance, are significant. I happen to be very sympathetic to this idea. In fact, it accommodates, either directly or as part of some more elaborate explanation, so wide a range of sensory phenomena, that it is undoubtedly true, at least to a close approximation. But this idea is not included in the list of six primary assumptions on page 140.

In fact, the author tells me: “As I mentioned as early as page 3 (and countless times thereafter), we are not using classical information theory in the development of perceptual theory, but rather we calculate differences in Boltzmann entropies, which are measured, arbitrarily, in bits” (personal communication, 22nd August 1994). This is not something that the reader could learn from a perusal of the book. Apart from an introductory chapter on thermodynamics, Boltzmann entropy is mentioned on the aforesaid page 3, on half of page 144, and then not again until two chapters from the end. I found the introductory chapter on thermodynamics less than adequate. In particular, it makes no attempt to relate Boltzmann entropy to perception, and I will again endeavour to supply the explanation that is lacking.

Envisage a visual experiment with grating stimuli. Suppose the area of the grating to be divided notionally into a fine grid of cells. If it helps, these cells could be thought of as corresponding to the mosaic of receptors in the retina or to the array of ganglion cells; but, really, the grid is simply a mathematical device to support calculation. Suppose $p_i$ is the probability that there will be a quantum emitted from cell $i$ and absorbed in the retinal receptors (quanta emitted, but not absorbed, are of no interest) within some short interval of time. The sum $-\sum_i p_i \ln p_i$ is the Boltzmann entropy of the stimulus. The $p_i$ will vary across the face of the grating, being higher in the peaks and lower in the troughs. Subject to a given space-average luminance ($\sum_i p_i$ fixed), the entropy is a maximum when all the $p_i$ are equal (that is, when the contrast is zero and the stimulus a ganzfeld).
During passage through the optic media, the visual stimulus suffers a small amount of scatter and also some diffraction at the pupil. The distribution of quanta is randomly rearranged, locally, with some reduction of the peaks and filling-in of the troughs. That makes the \( p_i \) more nearly equal, increasing the entropy. The perturbations of cellular potentials created by the absorption of quanta in the receptors suffer a similar random relocation in the course of lateral transmission from receptors to ganglion cells, leading to a further increase in entropy. So, the second law of thermodynamics transposes into the theorem that there can be no reduction in entropy (or increase in contrast) in the transmission of a stimulus through the sensory pathway. But I question whether this theorem is of much use to perceptual psychologists because it omits any consideration of the interaction between excitation and inhibition within receptive fields, and that, surely, is the most important component of sensory analysis.

The author, however, calculates entropy, not for specific stimuli, but at a fixed point in the sensory system. 'Quantum receptors' pass an averaged input—averaged over the stimulus input, but not over the added noise—to some higher level where the variability of those averaged inputs \( S \) in formula (4) is calculated. As incoming stimuli do 'work' on the system, there can be no analogue to the second law of thermodynamics, and the connection with Boltzmann entropy is no more than formal after all. But it is thereby understandable that an entropy-like statistic should sometimes decrease with time \([1/t]\) in formula (4), a property which is used to model adaptation.

The same mathematical property is used a second time to model the relation between the duration of a flash of light and its threshold luminance. This implies that the averaging of quantal input to the receptors must begin afresh at the precise point in time at which the flash begins. This epoch is a parameter of the configuration of the stimuli, which is not otherwise signalled to the subject. The question of how the quantum receptors can know when to begin calculating a fresh average is not addressed.

I cannot commend this book to readers of *Perception*. Although the mathematical argument is quite correct, algebraically speaking, it is, at the same time, a collage of otherwise unrelated mathematical motifs. It does not transpose into any coherent theory of perception—indeed, each time I take the book up I seem to find yet further arbitrary assumptions hidden in the mathematics, and the assignment of arbitrary interpretations to mathematical formulae is as exceptionable as the making of arbitrary verbal claims—and the question of its concordance with the state of nature scarcely arises. Some of the arguments are psychologically valid, others are not. To pick out the valid from the invalid, one needs to be expert in information theory.

In his preface the author explains that the initial data from which this book has grown date from 1958. For a book which has evolved over so long a period of time, one would wish to see a broader and deeper coverage of psychophysics and of sensory discrimination. Signal detection theory and psychometric functions are entirely missing. That is a strategic error, because the problems encountered in treating those topics might have tuned the theoretical treatment up to a point of usefulness. If one is looking at threshold values only, a variety of theoretical accounts are possible. That variety is much reduced if it be required that one and the same theory shall accommodate psychometric functions and signal detection as well.

D R J Laming
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England

References
Kullback S, 1959 *Information Theory and Statistics* (New York: Wiley)
Marks L E, 1974 "On scales of sensation: prolegomena to any future psychophysics that will be able to come forth as science" *Perception & Psychophysics* 16 358–376
Shannon C E, 1948 "A mathematical theory of communication" *Bell System Technical Journal* 27 379–423 and 623–656
Shannon C E, 1949 "Communication in the presence of noise" *Proceedings of the Institute of Radio Engineers* 37 10–21
Teghtsoonian R, 1971 "On the exponent in Stevens' Law and the constant in Ekman's Law" *Psychological Review* 78 71–80