Resurrecting Information Theory

Information, Sensation, and Perception

The heyday of information theory in psychology occurred sometime during the 1950s. Inspired by the seminal work of Shannon (1948), hundreds of articles during this period attempted to show that the perceptual and cognitive capacities of organisms were limited by the information content of the stimulus, and by the ability of the organism to transmit that information. The most widely known articles concentrated on absolute identification (e.g., Garner & Hake, 1951; Pollack, 1952) and reaction time (e.g., Hick, 1952; Hyman, 1953), but many other problems were attacked in a similar fashion (e.g., Attneave, 1954; Bendig, 1953; Miller, Heise, & Lichten, 1951). Comprehensive reviews of this work were written by Attneave (1959) and Garner (1962).

In 1956 George Miller published his classic article, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." Among other contributions, Miller showed that human memory is not limited by the amount of information in the message (i.e., as measured in bits), but rather by the number of chunks. In other words, recoding the message into larger chunks significantly increases the number of bits of information that can be recalled. Miller's article dealt what turned out to be a fatal blow to information-based theories of perceptual and cognitive capacity. Within a decade, the flow of articles had diminished to a trickle, and by the 1970s, information theory was mostly only of historical interest. For example, Laming's 1973 book, Mathematical Psychology, devotes fewer than 10 pages to information theory, and much of this focuses on its inadequacies.

Within the last few years, however, information-theoretic approaches have begun to reappear in the psychological literature, although the current applications are quite different from those of the 1950s. For example, one exciting new use of information theory is to solve the problem of inferring an unknown probability distribution when one is given only some summary statistics. An inference procedure, called maximum entropy inference, solves this problem by constructing the probability distribution that maximizes the information (or, equivalently, the entropy) of the distribution, subject to the constraints imposed by the known summary statistics (Jaynes, 1957; Shore & Johnson, 1980; Tribus, 1969). Recently, maximum entropy inference has been used to (a) account for learning in the mammalian hippocampus (Levy, 1989), (b) justify Shepard's (1987) theory of stimulus generalization (Myung & Shepard, 1994), (c) motivate decision-bound and exemplar models of category learning (Myung, 1994), (d) justify the privileged status that humans accord to basic-level categories (Corter & Gluck, 1992), and (e) construct normative models of human reasoning (Myers & Osherson, 1992).

Norwich's book is another example of a recent publication that borrows
heavily from information theory. His goals are lofty. Using an information-theoretic approach, he tries to develop a unified theory of sensation and perception by starting with a single fundamental equation of beguiling simplicity. Specifically, the entire book develops, tests, and discusses the equation \( F = kH \), where \( F \) is "a perceptually related variable such as subjective magnitude or impulse frequency in a sensory neuron" (p. 133), \( k \) is a positive constant, and \( H \) is the entropy or information of the stimulus.

Norwich's enthusiasm for the equation \( F = kH \) is apparent. For example, the back cover proclaims that his book shows "how nearly all of the empirical laws of sensory science discovered in the last 130 years can be derived from this equation." He also has little regard for the accomplishments of those in the field of theoretical psychophysics. For example, he says, "I chose the field of sensory science as a proving ground because this field is replete with empirical equations . . . but which are without any known theoretical basis" (p. xv). Of the previous attempts to unify various psychophysical laws, he says that "hundreds, if not thousands, of pages have been published in this endeavor. Yet the unification emerges easily from the entropy equation" (p. 151). With quotes such as these, Norwich's personality is never far from the reader's mind.

The first half of the book (chapters 1–8) surveys material needed to understand and appreciate the equation \( F = kH \). This includes a review of empirical laws in psychophysics, a brief description of information theory, and selected surveys of applications of information theory in psychology and physics. Chapters 9 through 14 develop, describe, and test the \( F = kH \) equation. The last three chapters discuss philosophical issues related to the equation and speculate on new applications.

The basic idea underlying the equation \( F = kH \) is that sensory receptors signal their uncertainty about the value of a stimulus, rather than signal the magnitude or frequency of a stimulus. Thus, rods and cones do not signal the brain about the amount of light they have received. They signal how uncertain they are about the mean luminance of the light they received. This is done in the following way. First, each receptor cell takes repeated samples of the stimulus. Presumably, with visual stimuli, each sample value will be a count of the number of photons arriving at the receptor during some fixed period of time. Suppose that, for a given stimulus, the mean number of photons arriving at the receptor is \( \mu_s \) and the variance is \( \sigma_s^2 \). By the central limit theorem, the sample mean of these values will have an approximate normal distribution with mean \( \mu_s \) and variance \( \sigma_s^2/m \), where \( m \) is the total number of samples. Second, Gaussian noise is added to the distribution of sample means. Norwich argues that this noise is really a reference signal generated by the sensory system in order to calibrate the stimulus entropy. He assumes it has mean 0 and variance \( \sigma_r^2 \), so when added to the distribution of the mean stimulus intensity, the result is a normal distribution with mean \( \mu_s \) and variance \( \sigma_s^2/m + \sigma_r^2 \). Third, the receptor is assumed to compute the entropy of this normal distribution (denoted by \( H_{s+r} \)). The absolute information in the stimulus, however, can be computed
only by subtracting the entropy in the reference signal (denoted by $H_r$). So in the final step, the receptor computes the difference $H = H_{s+r} - H_r$. This value is then relayed to sensory cortex and the magnitude of the subject’s ultimate sensory experience is assumed to be proportional to $H$.

As it stands, this scenario has a fundamental shortcoming. The entropy of a normal distribution increases with the population variance, but is unaffected by the mean. Thus, if mean luminance were to increase without a concomitant increase in the variance, Norwich’s hypothetical receptor would not change its output. In other words, perceived brightness would be unaffected by stimulus intensity. To circumvent this problem, Norwich assumes that the variance of the stimulus sample values, $\sigma_s^2$, is proportional to some power function of the stimulus intensity (i.e., so that $\sigma_s^2 = \alpha I^n$, for some constants $\alpha$ and $n$).

Norwich develops this model in chapter 9 and then devotes the next three chapters to a series of demonstrations to show that the model can account for, or nearly account for, a number of well-known psychophysical results. Included in this list are Fechner’s and Stevens’s laws, the classic adaptation curves, the near-miss to Weber’s law (although the theory has difficulty with the terminal rise in the Weber fraction), and a number of simple reaction time results. This is an impressive list that would challenge the best of modern psychophysical theories.

A closer examination reveals some potential problems with these derivations, however. For example, when deriving Stevens’s law, Norwich argues that the Stevens exponent should equal the exponent $n$ from his $\sigma_s^2 = \alpha I^n$ assumption. Consider the case of vision. Stevens exponents on brightness are about 0.33 (Stevens, 1962). Thus, if Norwich is correct, the variance of the stimulus samples received by the visual receptors should equal $\sigma_s^2 = \alpha I^{0.33}$, where $I$ is light intensity. However, it is well known that the number of photons emitted by a light source of constant intensity has a Poisson distribution (e.g., Wyszecki & Stiles, 1967). Thus, the number of photons arriving at a receptor during the fixed sampling period should also be Poisson distributed. In a Poisson distribution, the variance equals the mean, so the distribution of the sample mean would be approximately normal with mean $\mu_s = I$ and variance $\sigma_s^2 = I/m$. In other words, according to Norwich’s model, the Stevens exponent for brightness should be 1.0, not 0.33.

Norwich’s insistence that absolute entropy is computed by the receptor creates a number of special difficulties. First, it leaves little for the brain to do. According to Norwich, the end result of all processes after the receptor is to only multiply the signal by a positive constant (the $k$ in $F = kH$). One consequence of this assumption is that Norwich ignores all decision processes. In fact, he explicitly says that he “shall lump together all methods of measuring subjective magnitude and refer to them as just that: subjective magnitude” (p. 152). One of the great lessons of signal detection theory, however, is that decision processes are ubiquitous; they play an important role in even the simplest psychophysical tasks (e.g., Green & Swets, 1966). Further, it is
a mistake to assume that decision processes in all psychophysical tasks can be modeled by a simple multiplicative constant. For example, subjects apparently choose a much lower criterion for saying "yes" in a detection experiment when the stimuli are olfactory than when they are visual (e.g., Engen, 1972). As a result, when discriminabilities are equal, there will be more false alarms and hits with olfactory stimuli, but more misses and correct rejections with visual stimuli. It is unclear how the equation $F = kH$ could account for such differences.

The problems are equally serious at the receptor level. First, of course, is the problem that arises because Norwich assumes that the receptors perform a good deal of sophisticated computation, without specifying how the computations are performed. Much is now known about the biochemistry, physiology, and psychophysics of receptor events, especially in vision (e.g., Palczewski, 1994; Schoenlein, Peteanu, Mathies, & Shank, 1991; Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990). Currently, there is little or no evidence that receptors compute entropies of hypothetical probability distributions.

Second, consider the assumption that when the receptor computes the entropy of the distribution of stimulus samples, it assumes a normal distribution with variance $\sigma^2/\tau$ (the distribution prescribed by the central limit theorem). On any single stimulus exposure, a receptor receives $m$ stimulus samples, from which it can compute one mean. Norwich assumes that from this single number the receptor somehow knows that over many replications of this experiment, the distribution of all possible sample means will be approximately normal and that the variance will be $m$ times smaller than the variance of the sample. This kind of knowledge would require a prodigious memory. Certainly, there is no current evidence that sensory receptors have the capacity to anticipate the central limit theorem.

The problems at the receptor level could be solved by assuming that the entropies are computed and subtracted somewhere in the sensory cortex. Norwich resists this alternative because he interprets the $F$ in the $F = kH$ equation flexibly. In most applications, $F$ is a perceptual variable. But in other cases, $F$ is a measure of physiological activity. For example, Norwich sometimes interprets $F$ as the number of impulses recorded from some neuron early in the sensory system (e.g., as on p. 163). This interpretation is valid only if all entropy calculations are completed by earlier (i.e., more peripheral) sensory processes.

Much of Norwich's book seems new and exotic. For example, he uses Heisenberg's Uncertainty Principle, in part, to derive Miller's (1956) magical number seven (pp. 262–264). Elsewhere, he claims that "all perceptual units actively generate signals of the type which they perceive. . . . the auditory perceptual unit must generate audible sound if it is to perceive sound. . . . the visual perceptual unit must generate visible light if it is to perceive light signals" (p. 289). Despite these surprises, however, Norwich is, at heart, a classical psychophysicist. First, as mentioned above, he ignores decision processes. Second, he believes in a fixed perceptual threshold. As evidence of
this, he says "a threshold for sensation is the stimulus of least intensity which one can detect" (p. 31) and "it is well known that when I is less than some value, I\text{\textsubscript{thresh}}, . . . a stimulus is imperceptible" (p. 148).

Third, he minimizes the role of trial-by-trial variability in the percept. A casual reading, however, gives one the opposite impression. For example, several chapters are devoted to probability distributions, and there is extended discussion about receptors as quantum detectors. However, according to Norwich, once past the receptors, the stochastic nature of the signal disappears. This assumption is apparent in statements such as the following:

The receptors are quantum receptors and, therefore, detect quantum fluctuations. However, the brain, which receives the report of sensory receptors, is not made aware of quantum fluctuations. The message the brain receives concerns the level of the steady, macroscopic stimulus, which is the mean of the signals detected by the receptors. The fluctuating environment of the sensory receptor is translated, presumably by a process of averaging, into a smooth report received at the brain. Since the message of the sensory receptors is usually relayed to the brain by means of a neural frequency code, it is expected that this frequency will also, usually, encode a smoothed or averaged signal. (p. 136)

All three of these positions—that decision processes are unimportant, that a fixed threshold exists, and that there is no sensory and perceptual noise in the brain—are counter to the principles of signal detection theory (e.g., Green & Swets, 1966). Signal detection theory revolutionized psychophysics, and it continues to dominate the field today. If anything, its influence is expanding, as the ideas of signal detection theory are applied to tasks that are more and more complex (e.g., Ashby, 1992). Unfortunately, therefore, many current psychophysicists and cognitive scientists will find their views to be incompatible with the basic tenets of Norwich's book.

F. Gregory Ashby
Department of Psychology
University of California at Santa Barbara
Santa Barbara, CA 93106
E-mail: ashby@psych.ucsb.edu

References


