The magical number seven: Making a "bit" of "sense"

KENNETH H. NORWICH

Department of Physiology and Institute of Biomedical Engineering
University of Toronto, Toronto, Ontario M5S 1A8, Canada

The maximum number of categories distinguishable in making an absolute judgment was estimated by Miller to be "seven plus or minus two," corresponding to about 3 bits of information transmitted per stimulus. Later work extended this range to include at least 2 to 4 bits of information, which reached 16 categories. In contrast, the number of distinguishable differences between two stimuli is in the order of 100. Why is this so? It is shown here that an answer to these questions can be obtained by constructing an *entropy* function, H_s , which is a measure of the uncertainty of a subject (or a sensory receptor) as it perceives the magnitude of an applied stimulus. Using this function, it is demonstrated that the ubiquitous 3 bits of information per stimulus can be approximated from the expression $\log_2 \sqrt{\tau_2/\tau_1}$, where τ_1 and τ_2 are known time constants. The same entropy function can be used to derive various other psychophysical laws, such as the Weber-Fechner law, Stevens' law, and the Bunsen-Roscoe law.

When a human subject is presented stimuli selected at random from a set of discrete stimuli and required to select an appropriate response from a set of discrete responses, the mean information transmitted per stimulus under certain controlled circumstances has been found to be about 2.5 bits, corresponding to $2^{2.5}$, or about 7, categories (Miller, 1956). In this paper, we shall seek the origin of the "magical number" using a theoretical approach that differs somewhat from earlier ones.

The experimental paradigm for the "category" experiments is as follows. The stimulus is varied in only a single dimension; for example, if the stimulus is a tone, only the intensity is varied while the frequency is held constant. The stimulus is presented as a step function; that is, it is applied and held constant. The continuum of stimuli (say the intensity of tones) is divided into Q categories (e.g., intensities between the limits S and $S + \Delta S$ to category 1, between the limits $S + \Delta S$ and $S + 2\Delta S$ to category 2, etc., until category Q). The categories in this paradigm are wide, so that no problem of discrimination between categories exists. The subject is presented a number of stimuli which are selected randomly and is required to identify an appropriate response category. The subject will sometimes succeed and sometimes fail to make the correct identification or categorical judgment. The investigator can tabulate the results and

calculate the information correctly transmitted to the subject in bits per stimulus by the method described by Garner and Hake (1951). In experiments of this nature, we shall consider primarily those cases in which the applied stimuli span nearly the total normal physiological range (when this range can be identified). For example, pitch should span nearly the range from highest to lowest perceptible pitch.

It seemed rather surprising, in Miller's early survey, that the information transmitted per stimulus fell in the range of 2.5 bits, corresponding to 7 plus or minus 2 categories, for an exceedingly large number of types of stimulus. And, indeed, a later compendium of results prepared by Harper (1972) extended the range at least from 2.3 to 3.9 bits. Even this range of values is not rigid. Various factors can modify a subject's performance: degree of training, attention, "edge effects," and many others. Therefore, in considering a generalized perceptual system, and the performance of a general subject, one can really ask only such questions as: Why is the *range* of values about 2-4 bits per stimulus? And why is the *order of magnitude* about 3 bits per stimulus?

In marked contrast to the above, when a subject's ability to discriminate between nearly identical stimuli is tested over the normal physiological range, the number of distinctions he can make is relatively very large. For example, he can distinguish over 100 differences in the intensity of tone. But he can distinguish only about 7 categories in an absolute judgment experiment. Why is this so? Why is our ability to discriminate so much more acute than our ability to make absolute judgments?

In this theoretical examination, we shall see, with the aid of some mathematics, that the answers to all of the above questions, and a reasonable explanation

This work has been supported by a grant from the Natural Sciences and Engineering Research Council of Canada. I am grateful to Charles Cass for his aid in performing some of these calculations. I am most indebted to Dr. Sharon Abel of the Department of Otolaryngology, Mount Sinai Hospital, Toronto, and to Professor David Green of the Psychophysics Laboratory, Harvard University, for their guidance.

for a number of other psychophysical laws, may be associated with a seemingly unrelated phenomenon: the process of sensory adaptation.

The Entropic View of Sensory Perception

Very briefly, the entropic view of perception states that we can perceive only when we are uncertain; we cease to perceive when we achieve complete certainty (Norwich, 1977, 1978). Our level of uncertainty at any stage in the perceptual process can be measured using Shannon's measure of uncertainty or *entropy*. The difference between the uncertainty or entropy at the beginning of a perceptual act, $H_{\max}$, and at the end of the perceptual act, $H_{\min}$, gives a measure of the amount of information which has been transmitted during the perceptual act. Now, in a Garner-Hake experiment, one calculates just this quantity of information: the stimulus information, I_s , minus the stimulus equivocation, which we shall call H_s , equals the transmitted information I_t . If these two measures of transmitted information are the same, the difference $H_{\max} - H_{\min}$ will be equal to the difference $I_s - H_s$; and the maximum value of the difference $H_{\max} - H_{\min}$ will be equal to the maximum value of the difference $I_s - H_s$. This maximum is the logarithm of Miller's magical number. We shall calculate this maximum reduction in entropy both psychophysically, by the classical method of Garner and Hake, and physiologically, by following the variation in entropy as it drops from its maximum value at the moment of imposition of a stimulus (light, sound, color, etc.) to its lowest value, when further exposure to the stimulus will not impart any appreciable additional information to the subject.

One should realize that the act of division of a stimulus continuum into a number of discrete categories does not *produce* the information content of that continuum. Rather, it *calibrates* the continuum so that its information content may be measured. In the same way, if one divides the interval between two points into a number of equally spaced intervals, one does not, in the process, *produce* the separation between the points, but simply provides a means of measuring the separation. Increasing the number of equally spaced categories/length intervals is equivalent to diminishing the unit of measurement, but it does not affect the total information content/two-point separation.

The progressive acquisition of information: Psychophysical and physiological adaptation. Let us consider the process of acquisition of information first from a psychophysical and then from a physiological point of view. The psychophysical approach ($I_s - H_s$) is currently the only feasible approach to more complex stimuli such as hue, pitch, length of lines, etc. The physiological approach ($H_{\max} - H_{\min}$) can be used in the case of simpler stimuli, such as the

magnitude of a light, tone, force, etc., applied to a single, isolated sensory receptor. In the psychophysical case, the response to the stimuli will be replies of the human subject. In the physiological approach, the response will be the rate of transmission of impulses along the primary afferent neuron issuing from the sensory receptor. We shall have to translate both types of response into the same language, namely bits of information. We shall then see that the psychophysical and physiological approaches are complementary.

Suppose that we conduct a category-identification experiment in the following way. In the first instance a stimulus, such as a flash of light or a pulsed tone, is presented to a subject for a very brief period of time. The exposure time exceeds the threshold exposure time for a stimulus of that intensity, but not by much. The subject is required to identify the category of the stimulus in the usual way. The experiment is repeated many times, using stimuli of different intensities, but always of the same duration or exposure time. The responses are tabulated, and a stimulus-response matrix is compiled in the manner described by Garner and Hake (1951). The stimulus equivocation, H_s , is calculated in the usual manner. H_s will be quite large, because the subject will be very uncertain about the nature of the stimulus after such a brief exposure.

Now let us repeat the experiment with the stimulus applied for a longer period of time. A second value for the stimulus equivocation, H_s , is calculated, and so on. The experiment is repeated many times, with the duration of exposure of the subject to the stimulus increasing progressively. For a while, the value of H_s falls monotonically, because the subject makes fewer and fewer incorrect identifications as his duration of exposure to the stimulus lengthens. But, after a certain duration of exposure to the stimulus, the value of the equivocation, H_s , will fall no further. If we were to draw a graph of H_s against stimulus duration, we should probably obtain something like the graph shown in Figure 1. The graph does not fall to $H_s = 0$, since errors in identification will be made even when the stimulus is of infinite duration. Such a curve might be called a *psychophysical adaptation curve*. It describes the progressive decrease in uncertainty associated with progressively longer periods of exposure to a stimulus.

From the work of Garner and Hake, we write

$$H_s = I_s - I_t, \quad (1)$$

where I_s is the stimulus information, or entropy, and I_t the transmitted information. When τ , the duration of application of the stimulus, is very small, very little information is actually transferred to the subject. Under these conditions, I_t is very small and H_s is a maximum. Using Equation (1) we see that

Figure 1. Schematic diagram of a "psychophysical adaptation curve." τ is the duration of application of the stimulus. H_s , the entropy of equivocation, represents the uncertainty of the subject about the nature of the stimulus (hue, pitch, etc.). H_s is calculated by the methods of Garner and Hake (1951). Each point on the graph is calculated from an entire categorical judgment experiment consisting of many S-R pairs.

$$H_{s \max} = I_s. \quad (2)$$

When τ , the duration of application of the stimulus, becomes large, the information transmitted to the subject is maximal. That is, I_t is maximum when H_s is minimum. From Equation (1) we see that, under these conditions,

$$H_{s \min} = I_s - I_{t \max}. \quad (3)$$

Combining Equations 2 and 3, we obtain

$$I_{t \max} = H_{s \max} - H_{s \min}. \quad (4)$$

But $I_{t \max}$ (bits of information) is the maximum possible quantity of information transmitted in the categorical judgment experiment. Therefore, expressed in another way

$$\log_2(\text{magical number}) = \underset{\text{bits}}{H_{s \max} - H_{s \min}}. \quad (5)$$

Equations 4 and 5 show us that the information capacity of a subject in performing a categorical judgment task can be obtained by subtracting the lowest from the highest ordinate on a psychophysical adaptation curve.

Let us now consider the progressive acquisition of information from the physiological point of view. Consider an isolated sensory unit consisting of a sensory receptor (light receptor, stretch receptor, etc.) with its primary afferent sensory neuron. Let us apply a steady stimulus of long duration to the

receptor, and record the firing rate of the sensory neuron (impulses per second).

The neuron will fire initially at maximum rate, progressively slowing its firing rate until it reaches an equilibrium level. The graph of firing rate against time describes the *physiological adaptation curve* of this sensory unit (Figure 2). The physiological adaptation curve has the same general shape as the psychophysical adaptation curve shown in Figure 1. It therefore seems reasonable to associate the firing rate of the neuron, F , with the entropy or stimulus equivocation, H_s . We write, then, by hypothesis (Norwich, 1977),

$$F = kH_s, \quad (6)$$

where k is a constant that is greater than zero. Applying the same reasoning that we followed with the psychophysical adaptation curve, we write for the physiological curve

$$I_{t \max} = H_{s \max} - H_{s \min} \quad (4)$$

$$= (F_{\max} - F_{\min})/k. \quad (7)$$

That is, the maximum transmitted information, $I_{t \max}$ (which is the "magical" number of bits of information transmitted), is proportional to the difference between the maximum and minimum firing rates of the sensory neuron.

Now, in the case of the psychophysical adaptation curve, we determined the value of $I_{t \max}$ using Equation 4 from the quantities $H_{s \max}$ and $H_{s \min}$, which were determined by the methods of Garner and Hake. In the case of the physiological adaptation curve, we cannot determine $H_{s \max}$ and $H_{s \min}$ directly. We can, however, measure $F_{\max}$ and $F_{\min}$. Therefore, by

Figure 2. Schematic diagram of a "physiological adaptation curve." F is the number of impulses per second transmitted by the primary afferent neuron. F decreases from its maximum value, $F_{\max}$, to its steady rate, $F_{\min}$, while the receptor received a constant, uniform stimulus.

reference to Equation 7, we require only the value of the constant k in order to determine the "magical" number of bits of information transmitted per stimulus. In order to obtain a value for k , we shall construct a simple information model of the process whereby a receptor detects a steady signal. That is, we shall derive the mathematical form of the function H_s , which will then be inserted into the right-hand side of Equation 6. By a process of curve-fitting we shall then be able to find a value for the constant k .

The mathematical form of the entropy function, H_s , for a sensory receptor. This section is based upon the concepts of information theory, some of which are reviewed in Appendix B. This account is essentially a review of the arguments presented by Norwich (1977).

We understand that the acquisition of information is never instantaneous, whether we deal psychophysically with a subject to whom a stimulus is presented or physiologically with a sensory receptor to which a stimulus is applied. Although the stimulus is a steady signal, a tone, light, etc., it consists nonetheless of a spectrum of intensities. The mean of this spectrum of intensities is constant (for a steady signal) and is, of course, a measure of the central or prevailing effect of the stimulus. The signal detector (receptor) samples the signal progressively, and eventually distinguishes the value of the mean. We can assign a probability density function, $p(I)$, to the spectrum of intensities, such that $p(I) dI$ is the probability that a sample of the signal will yield an intensity which lies in the interval I to $I + dI$. In the case of certain types of signal or stimulus, we do, indeed, know the explicit mathematical form of the function $p(I)$. But, in general, we do not. It could be a Gaussian or normal distribution, a lognormal distribution, Poisson distribution, etc. But if we consider the mean value of groups of n samples taken by the receptor of its signal, we know from the central limit theorem of statistics that, regardless of the nature of the original signal, as n becomes large, these means will be distributed in a Gaussian distribution. If σ^2 is the variance of the original distribution, $p(I)$, then σ^2/n is the variance of the distribution of means. Since it is the mean, or central, tendency which the receptor "seeks," it is the probability density function for this mean with its variance σ^2/n which we should consider.

In order to complete the model of the process of progressive acquisition of information, it is necessary to make a few simplifying assumptions. It will be assumed that the noise signal, which is always of great importance when continuous signals are being sampled, is also Gaussian and has the same mean value as the message signal. Let us represent the probability density function for noise by $p_N(I)$ and let the noise variance be N^2 . Let the probability density function for the signal together with the noise be $p_{SN}(I)$. That

is, $p_{SN}(I) dI$ is the probability that a sample of signal together with noise will have an intensity between I and $I + dI$. The function $p_{SN}(I)$ is obtained by means of a convolution integral, and can be shown to have variance $\sigma^2 + N^2$. That is, the variance of the signal that is a mixture of the message signal with the noise signal is equal to the sum of the component variances.

We define the entropy or uncertainty about the mean, at the receptor, by the expression

$$H_s = - \int p_{SN}(I) \log p_{SN}(I) dI + \int p_N(I) \log p_N(I) dI. \quad (8)$$

Since both p_{SN} and p_N are known Gaussian or normal distributions, the value of H_s can be obtained. This integration is given explicitly in Appendix B. After carrying out the integration, we obtain

$$H_s \text{ (natural units)} = \frac{1}{2} \log_e \left(1 + \frac{\sigma^2/n}{N^2} \right) \quad (9)$$

or

$$H_s \text{ (bits)} = \frac{1}{2} \log_e \left(1 + \frac{\sigma^2/n}{N^2} \right) \cdot \frac{1}{\log_e 2}. \quad (10)$$

By definition, sample size, $n \geq 1$.

Finally, it is necessary to relate the number, n , of samples acquired by the receptor to the time, τ , of exposure to the stimulus. In the absence of any definitive information on this subject, we shall assume that the receptor samples its environment at regular intervals, so that

$$n = \gamma \tau, \quad (11)$$

where γ is a constant greater than zero.

Introducing Equation 11 into Equation 10, we obtain

$$H_s \text{ (bits)} = \frac{1}{2} \log_e \left(1 + \frac{\sigma^2/\gamma\tau}{N^2} \right) \cdot \frac{1}{\log_e 2}. \quad (12)$$

Setting

$$B = \sigma^2/\gamma N^2, \quad (13)$$

we have

$$H_s \text{ (bits)} = \frac{1}{2} \log_e \left(1 + B/\tau \right) \cdot \frac{1}{\log_e 2}. \quad (14)$$

B is constant for a signal with constant mean value. Combining Equations 6 and 14, we have the result we were seeking:

$$F = \frac{1}{2}k \log_e(1 + B/\tau) \cdot \frac{1}{\log_e 2}. \quad (15)$$

F is the firing rate of the sensory neuron while a steady signal is being applied. τ is time. k is a constant whose value we do not yet know.

We are now in a position to complete the analysis of the physiological adaptation curve.

Analysis of Experimental Data from Physiological Adaptation Curves

The order of procedures is as follows: (1) Curve-fit Equation 15 to the measured adaptation curves, and in this way evaluate the parameter k . (We obtain the value of B in the process.) (2) Estimate values for $F_{\max}$ and $F_{\min}$, the highest and lowest firing rates observed in the adaptation curve. (3) Calculate the information transmitted to the receptor, $I_{t \max}$, using Equation 7 (Norwich, 1979). We can then compare the maximum transmissible information in a sensory adaptation process to the maximum transmissible information ("magical" number of bits) in a psychophysical adaptation curve.

Since there have not been many physiological adaptation curves from human subjects published, it is necessary to use data from animal experiments. The data to be presented have been taken from two papers which are readily available to most readers, should they wish to refer to them. The data have been digitized from graphs, and the approximate numbers have been entered in Tables 1 and 2.

Steps 1 to 3 above were carried out, and the results have been entered in Tables 1 and 2. The curve-fitting was done using a Hewlett-Packard 9825A desk calculator, by a method that is described in Appendix A. The fitted values for parameters k and B have been given in the tables so that the accuracy of all results can be checked.

Table 1 contains the data of Matthews (1931), who applied forces to a single muscle spindle in a small muscle in the middle toe of a frog and recorded the transmission rate of impulses in the peroneal nerve. Table 2 contains the data of Galambos and Davis (1943), who applied an acoustical stimulus to the intact ear of a cat and recorded the rate of transmission of impulses in a single fiber in the auditory nerve.

From the analysis of Matthews' data, it is seen that when the stretch receptor was stimulated by 5-, 2-, and 1-g forces, the transmitted information for the frog's nerve was 2.55, 2.91, and 2.87 bits per stimulus, respectively. The equivalent number of categories was 5.86, 7.52, and 7.31, respectively. From the analysis of the data of Galambos and Davis, the auditory nerve fiber of the cat was seen to transmit 3.84 bits per stimulus, corresponding to 14.32 categories.

A similar analysis was carried out using the data of Hartline and Graham (1932) for a single ommatidium from the eye of the *Limulus* crab. The transmitted information was only .38 bits, or $2^{.38} = 1.30$ categories. This figure, if correct, would allow the *Lim-*

Table 1
Data of B. H. C. Matthews, Figure 15A

Time	Frequency					
	(a) Force of 5 g		(b) Force of 2 g		(c) Force of 1 g	
	Measured	Fitted	Measured	Fitted	Measured	Fitted
.50	95	95.6	65.0	65.1	45.0	45.4
1.00	86	83.9	57.0	56.5	41.0	40.0
2.00	74	72.3	49.0	48.1	36.0	34.6
3.00	65	65.6	42.0	43.2	31.0	31.5
4.00	60	60.8	38.0	39.8	29.0	29.3
5.00	57	57.1	35.0	37.1	27.0	27.6
6.00	54	54.2	33.0	35.0	26.0	26.2
7.00	50	51.7	32.0	33.2	25.0	25.1
8.00	48	49.5	31.0	31.7	24.0	24.1
9.00	47	47.6	30.0	30.3	23.0	23.2
10.00	46	46.0	30.0	29.1	22.0	22.4
11.00	45	44.6	30.0	28.1	21.5	21.7
12.00	43	43.1	29.5	27.1	21.0	21.1
13.00	43	41.8	29.5	26.2	21.0	20.5
14.00	42	40.7	29.0	25.4	20.5	20.0

Note—Stimulus = force; receptor = stretch; animal = frog. Measured frequency is given in impulses per second; fitted frequency = kH_s . (a) $F_{\max} = 100$, $F_{\min} = 40$; $k = 23.5$; $B = 140.5$. Transmitted information = $(F_{\max} - F_{\min})/k = 100-40/23.50 = 2.55$ bits. Magical number = $2^{2.55} = 5.86$ categories. (b) $F_{\max} = 75$, $F_{\min} = 25$; $k = 17.19$; $B = 94.73$. Transmitted information = $(F_{\max} - F_{\min})/k = 75-25/17.19 = 2.91$ bits. Magical number = $2^{2.91} = 7.52$ categories. (c) $F_{\max} = 50$, $F_{\min} = 19$; $k = 10.81$; $B = 167.1$. Transmitted information = $(F_{\max} - F_{\min})/k = 50-19/10.81 = 31/10.81 = 2.87$ bits. Magical number = $2^{2.87} = 7.31$ categories. In entering the values for $F_{\max}$ and $F_{\min}$ in Tables 1 and 2, some license has been taken. The values for $F_{\min}$ are slightly lower than the lowest measured frequency; this value is an estimate of the steady discharge rate toward which the values were asymptotically tending. Similarly, the values for $F_{\max}$ are slightly larger than the highest measured frequency.

Table 2
Data of R. Galambos and H. Davis, Figure 3

Time	Frequency	
	Measured	Fitted
.025	320	306.0
.050	180	264.0
.100	140	223.0
.150	150	200.0
.200	151	183.0
.250	149	170.0
.300	120	160.0
.350	125	152.0
.400	110	144.0
.450	127	138.0
.500	148	132.0
.550	120	127.0
.600	102	123.0
.650	102	118.0
.700	130	115.0
.750	100	111.0
.850	102	105.0
.900	122	102.0
.950	102	99.3
1.000	80	96.8
1.050	102	94.5
1.100	123	92.3
1.150	123	90.2
1.200	68	88.2
1.250	110	86.3
1.300	81	84.5
1.350	102	82.8
1.400	81	81.2
1.450	101	79.7
1.500	103	78.2
1.550	55	76.7
1.600	98	75.4
1.650	102	74.1
1.700	102	72.8
1.750	118	71.6
1.800	80	70.4
1.900	80	68.2

Note—Stimulus = tone; receptor = ear; animal = cat. Measured frequency is given in impulses per second; fitted frequency, kH_s , $F_{max} = 400$, $F_{min} = 80$; $k = 83.39$; $B = 4.00$. Transmitted information = $(F_{max} - F_{min})/k = 400 - 80/83.39 = 3.84$ bits. Magical number = $2^{3.84} = 14.32$ categories.

ulus crab only very restricted capability for identifying different intensities of illumination. It does not have any bearing, however, on the sensitivity of the omatidium to light stimuli.

We can see that the maximum transmissible information coded by a physiological adaptation curve, and measured by curve-fitting the function H_s to experimental adaptation curves, is of the same order of magnitude as the maximum transmissible information coded by a psychophysical adaptation curve and measured by the method of Garner and Hake.

There are three observations which should be made at this time:

(1) Because of the asymptotic approach of the adaptation curve to its lowest, or plateau, level, it is not necessary to apply a stimulus for an infinite

period of time in order to make a reasonable estimate of the plateau level. A stimulus of finite duration will very soon produce a close estimate of $I_{t \max}$.

(2) In order that the entropy H_s be allowed its maximum excursion ("magical" number of bits), it is necessary that $H_{s \max} - H_{s \min}$ take on its largest possible value. In the psychophysical case, this means that the stimuli in a categorical judgment experiment must span a "large" range of values; in the physiological case, this means that the signal applied to the sensory receptor must be "large." The magnitude of this range of stimuli and sensory signals can be calculated for a number of cases, but this calculation exceeds the scope of the present paper.

(3) In order to compare the results of rather simple, single-receptor experiments performed on animals with the rather sophisticated, absolute judgment experiments using human beings, we envisage, by way of simplification, all receptor-neuron units operating in parallel, divested of their elaborate collateral interconnections (Figure 3). Thus, all receptors working together are taken to produce a simple amplification of the response of a single receptor.

In the process of deriving the function H_s (Equation 10), we introduced several simplifying assumptions. In order to show that this function, H_s , has not been obtained in an ad hoc manner, the scope and utility of this function will now be demonstrated.

The Weber-Fechner Law as a Special Case of the Entropy Function, H_s

Equation 9 relates receptor uncertainty, H_s , to signal variance. We can relate signal variance, σ^2 , to signal mean, μ , using the empirical equation

$$\sigma^2 = b^2 \mu^{2m} \quad b, m \text{ constants} > 0. \quad (16)$$

This equation states simply that signals of higher intensity are associated with greater variances. It is obviously true for distributions such as the Poisson

Figure 3. Simplified physiological model for expressing the maximum transmissible information in a categorical judgment. Multiple, parallel receptor-neuron units carry essentially the same message to the brain. Collateral fibers are ignored.

distribution in which variance and mean are equal. Introducing Equation 16 into Equation 9,

$$H_s \text{ (natural units)} = \frac{1}{2} \log_e \left(1 + \frac{b^2 \mu^{2m}}{nN^2} \right). \quad (17)$$

We shall not require any further assumptions. The function H_s as described by Equation 17 is capable of explaining and predicting a great number of psychophysical laws and phenomena. We shall, however, examine it for high and low values of $b^2 \mu^{2m}/nN^2$.

For signals of high intensity (μ large) or with lower noise variance (N^2 small),

$$1 \ll b^2 \mu^{2m}/nN^2 \quad (18)$$

or equivalently, using Equations 11, 13, and 16,

$$1 \ll B/\tau. \quad (18a)$$

Equation 17 now simplifies to the form

$$H_s \doteq \frac{1}{2} \log_e \left(\frac{b^2 \mu^{2m}}{nN^2} \right) \quad (19)$$

or

$$\begin{aligned} H_s &= m \log_e \mu + \frac{1}{2} \log_e (b^2/nN^2) \\ &= m \log_e \mu + \text{a constant}. \end{aligned} \quad (20)$$

Equation 20 is a statement of the Weber-Fechner law. It gives the receptor uncertainty as a linear function of the logarithm of the mean stimulus intensity, μ . The entropy or uncertainty decrement, $H_s - H_{s \min}$, parallels the sensation or sensory effect produced by the stimulus and measures the information transmitted by the perceptual process (Equation 4). Multiplying both sides of Equation 20 by k , and introducing Equation 6,

$$F = km \log_e \mu + \text{a constant}, \quad (21)$$

which relates the firing rate, F , to the logarithm of the mean stimulus intensity. Equation 21 is a statement of the neurophysiological version of the Weber-Fechner law.

Stevens' Law as a Special Case of the Entropy Function, H_s

For signals of lower intensity (μ smaller), or with higher noise variance (N^2 large),

$$b^2 \mu^{2m}/nN^2 < 1. \quad (22)$$

$$H_s = \frac{1}{2} \log_e (1 + z),$$

where $z = b^2 \mu^{2m}/nN^2 < 1$, which can be expanded in a Taylor series,

$$H_s = \frac{1}{2} \log_e (1 + z) = \frac{1}{2} (z - \frac{1}{2} z^2 + \dots).$$

Terms of the second order and higher may be dropped due to Inequality 22. That is,

$$H_s \doteq \frac{1}{2} (b^2/nN^2) \mu^{2m} \quad (23)$$

or

$$H_s = \text{constant} \times \mu^{2m}. \quad (23)$$

Equation 23 is a statement of Stevens' law. It gives the receptor uncertainty as a power function of the mean stimulus intensity, μ . Again, the entropy of uncertainty decrement, $H_s - H_{s \min}$, parallels the sensory effect produced by the stimulus and measures the information transmitted by the perceptual process (Equation 4). Multiplying both sides of Equation 23 by k and applying Equation 6, we have

$$F = \text{constant} \times \mu^{2m}. \quad (24)$$

(Of course, the spontaneous firing rate adds another constant to the right-hand side of Equation 24). Equation 24 is a statement of the neurophysiological version of Stevens' law.

Prediction: When the stimulus intensity distribution is strictly Poisson ($m = \frac{1}{2}$), *for low intensities* the graph of $\log F$ (the logarithm of the rate by which the firing rate exceeds the resting or spontaneous firing rate) against $\log \mu$ should be a straight line with slope equal to unity (from Equation 24).

The Bunsen-Roscoe Law as a Special Case of the Entropy Function

This law states that if the product of intensity and duration in a light flash is constant, the visual effect is constant. We shall see that the law issues as an approximation from Equation 23. The normal distribution that has been adopted for the mean light intensity is not precisely correct because of the brevity of the light flash.

Incorporating Equation 11 into Equation 23,

$$H_s = \frac{1}{2} (b^2/N^2 \gamma) \cdot \frac{\mu^{2m}}{\tau}. \quad (25)$$

Light from an incandescent source emits photons such that the number captured by the receptor per unit time has a Poisson distribution. From the properties of the Poisson distribution, using Equation 16

$$2m = 1, \quad (26)$$

Equation 17 is now of the form

so that Equation 25 becomes

$$H_s = \frac{1}{2}(b^2/N^2\gamma) \cdot \frac{\mu}{\tau}. \quad (27)$$

Let us consider a flash with intensity μ_A . The earliest response of the receptor to the flash occurs at time $\tau = \tau_1$. The receptor entropy at time τ_1 is

$$H_{s_1} = \frac{1}{2}(b^2/N^2\gamma) \cdot \frac{\mu_A}{\tau_1}. \quad (28)$$

The flash is of duration $\Delta\tau_A$, and the receptor entropy at time

$$\tau_2 = \tau_1 + \Delta\tau_A \quad (29)$$

is given by

$$H_{s_2} = \frac{1}{2}(b^2/N^2\gamma) \cdot \frac{\mu_A}{\tau_2}. \quad (30)$$

The reduction in entropy, which is the information transmitted and the visual effect, is given by

$$\begin{aligned} H_{s_1} - H_{s_2} &= \frac{1}{2}(b^2/N^2\gamma)\mu_A \left(\frac{1}{\tau_1} - \frac{1}{\tau_2} \right) \\ &= \frac{1}{2}(b^2/N^2\gamma)\mu_A \left(\frac{\tau_2 - \tau_1}{\tau_1\tau_2} \right). \end{aligned} \quad (31)$$

Similarly, if we consider a flash with intensity μ_B , its visual effect is given by

$$H'_{s_1} - H'_{s_2} = \frac{1}{2}(b^2/N^2\gamma)\mu_B \left(\frac{\tau_3 - \tau_1}{\tau_1\tau_3} \right), \quad (32)$$

where the duration of the flash is $\Delta\tau_B$ and

$$\tau_3 = \tau_1 + \Delta\tau_B. \quad (33)$$

If the visual effect is the same for both flashes, we can equate the right-hand sides of Equations 31 and 32:

$$\mu_A \frac{\tau_2 - \tau_1}{\tau_2} = \mu_B \frac{\tau_3 - \tau_1}{\tau_3} \quad (34)$$

or

$$\frac{\mu_A \Delta\tau_A}{\tau_2} = \frac{\mu_B \Delta\tau_B}{\tau_3}. \quad (35)$$

For flashes that are nearly equal in duration ($\tau_2 \doteq \tau_3$), it can be shown that

$$\mu_A \Delta\tau_A = \mu_B \Delta\tau_B, \quad (36)$$

which is the Bunsen-Roscoe law.

Miller's Magical Number Estimated a priori from the Entropy Function

Introducing Equation 11 into Equation 19,

$$H_s = \frac{1}{2} \log_e \left(\frac{b^2 \mu^{2m}}{N^2 \gamma \tau} \right). \quad (37)$$

Let us now select two times τ_1 and τ_2 such that

$$H_{s \max} = \frac{1}{2} \log_e \left(\frac{b^2 \mu^{2m}}{N^2 \gamma} \right) - \frac{1}{2} \log_e \tau_1 \quad (38)$$

and

$$H_{s \min} = \frac{1}{2} \log_e \left(\frac{b^2 \mu^{2m}}{N^2 \gamma} \right) - \frac{1}{2} \log_e \tau_2 \quad (39)$$

$$\begin{aligned} H_{s \max} - H_{s \min} &= \frac{1}{2} \log_e \tau_2 - \frac{1}{2} \log_e \tau_1 \\ \text{natural units} &= \frac{1}{2} \log_e (\tau_2 / \tau_1) \\ &= \log_e \sqrt{\tau_2 / \tau_1}. \end{aligned} \quad (40)$$

The magical number of bits is then given by

$$H_{s \max} - H_{s \min} = \log_e \sqrt{\tau_2 / \tau_1} / \log_e 2 = \log_2 \sqrt{\tau_2 / \tau_1}. \quad (41)$$

bits

τ_1 is the time of the maximum response to a steady stimulus. τ_2 should be taken as the time when adaptation to the stimulus is, say, 90% complete. (We are using here Inequality 18, so that B/τ must stay large.) τ_2/τ_1 can be regarded as the ratio of two time constants.

For the data of Matthews given in Table 1, this estimate gives ($B/\tau > 9.5$)

magical number of bits

$$= \log_e \sqrt{10/.5} / \log_e 2 = 2.16 \text{ bits.}$$

For the data of Galambos and Davis given in Table 2, the estimate gives ($B/\tau > 2$)

magical number of bits

$$= \log_e \sqrt{2/.025} / \log_e 2 = 3.16 \text{ bits.}$$

Both estimates compare favorably with the more exact calculations given in the tables.

Reasoning with a priori knowledge of adaptation curves, we can provide an order of magnitude calculation of Miller's magical number. The maximum response to a steady stimulus is unlikely to occur before 20 msec after imposition of the stimulus. The period of 90% adaptation is not likely to exceed 10 sec. Therefore, the largest value expected for

Miller's number is about $\log_e \sqrt{10/.02} / \log_e 2$, which is about 4.5 bits of information.

Discrimination vs. Categorical Judgment

The ear, in a categorical judgment test, is capable of resolving only about 14 categories. But the ability of the ear to *discriminate* between two sounds is much finer. In the range of 30 to 100 dB, the ear can distinguish between two sounds differing by only 1 dB. This would span about 70 distinguishable differences as the stimulus changed between 30 and 100 dB (see, e.g., Mountcastle, 1980). How can one reconcile the 14 categories with the 70 distinguishable differences?

We shall demonstrate in the following argument that the maximum number of distinguishable differences is approximately equal to the number of impulses transmitted during the adaptation process of the sensory organ. Let us return to the adaptation curve for a single, sustained tone; the curve is described by the function $H_s(\tau)$. From Equation 6, we have

$$F = kH_s. \quad (6)$$

Suppose that F takes on its minimum value, $F_{\min}$, when τ has the value τ_{eq} , and that F takes on its maximum value, $F_{\max}$, when τ has the value $\tau_{\min}$. Both sides of Equation 6 may now be integrated with respect to time:

$$\int_{\tau_{\min}}^{\tau_{\text{eq}}} F d\tau = k \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau. \quad (42)$$

The integral on the left-hand side gives the total number of impulses that have been generated in the neuron as the firing rate dropped from $F_{\max}$ to $F_{\min}$. Designating a change or increment in the integral produced by a small change in the stimulus magnitude, by the operator Δ_μ ,

$$\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} F d\tau = k \Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau. \quad (43)$$

This equation relates the change in total number of impulses recorded during an adaptation curve, produced by a small change in stimulus magnitude, to the change in the integral of the sensory effect.

Dividing Equation 42 by Equation 43,¹

$$\frac{\int_{\tau_{\min}}^{\tau_{\text{eq}}} F d\tau}{\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} F d\tau} = \frac{\int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / (\tau_{\text{eq}} - \tau_{\min})}{\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / (\tau_{\text{eq}} - \tau_{\min})} \\ \doteq \frac{\text{mean entropy}}{\text{change in mean entropy}}. \quad (44)$$

Now the smallest possible change,

$$\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} F d\tau,$$

is simply one impulse, so that the corresponding value of

$$\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / (\tau_{\text{eq}} - \tau_{\min})$$

approximates the smallest perceptible amount of information transmitted. That is, in the limit, we have

$$\int_{\tau_{\min}}^{\tau_{\text{eq}}} F d\tau = \frac{\int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / (\tau_{\text{eq}} - \tau_{\min})}{\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / (\tau_{\text{eq}} - \tau_{\min})} \quad (45)$$

so that the integral on the left-hand side approximates the maximum number of distinguishable differences associated with a given stimulus level.

The integral on the left-hand side of Equation 45 can be evaluated numerically from the data for the cat's ear given in Table 2. The value is 206 impulses, with τ_{eq} taken as 1.9 sec. This value should be reduced by the resting or spontaneous noise level, which is about 80 impulses (Figure 3, Galambos & Davis, 1943). This still leaves $206 - 80 = 126$ distinguishable differences for the cat's neuron. The same neuron could resolve only 14.3 categories. We can now begin to understand why considerably less information is required to make a distinction between two stimuli than to actually identify a category.

Combining Equations 42 and 45 provides the relationship

$$\Delta_\mu \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau = \frac{1}{k}, \quad (46)$$

which indicates the manner in which k is related to the difference limen of information transmission (or sensory effect).

A brief history of the pursuit of the magical number: Sensory capacity and sensory memory

Rather a large number of studies have been made on the phenomenon of the apparent constancy of 2 - 4 bits of information transmitted per stimulus in categorical judgment experiments (corresponding very roughly to Miller's magical number seven plus or minus two). Siegel (1972) classified the hypotheses on the "magical number" which were then extant into two groups: the sensory capacity hypothesis, in which the sensory modalities limit the transmission of information; and the memory capacity hypothesis, in which the limitation on absolute judgment perfor-

mance is relegated to a more central mechanism. To these might be added a third, miscellaneous group, in which the "attention band" model of Luce, Green and Weber (1976) might be placed. The intensity resolution model of Durlach and Braida (1969) is a memory capacity model, and falls into the second group. Siegel's arguments showing the influence of memory are very convincing. He leaves no doubt whatever that either memory or the confusing effect of the number of intervening items severely limits performance in a categorical judgment experiment involving single tones. He has not, however, demonstrated that memory is the limiting factor in performance. Although his information transmission curves never leveled or reached a plateau, he used at most 9 categories, or just above 3 bits of stimulus information. The present work falls into the first of the groups, the sensory capacity group. Admittedly within the limitations of the data analyzed, it suggests that the sensory system does place an upper limit on the transmitted information. In the mammal, this upper limit may exceed 15 categories for tones, and is, therefore, in excess of the actual performance of the subject, suggesting that some information is lost more centrally.

The sources of loss of information in the central nervous system and the means for minimizing this loss have been the subjects of various papers. Perhaps the effect of practice and training is the chief of these factors. This has been discussed by Cuddy (1970) with respect to pitch identification, by Fulgosi and Bartolović (1971) with respect to visual stimuli, is mentioned by Harper (1972) with regard to odour, and is included in various monographs dealing with categorical judgments.

Because of the many factors which affect the calculation of the transmitted information, it must be stressed that any numerical value obtained will be encumbered with many qualifying statements. Investigators are generally most attentive to this. Even the method used to calculate the transmitted information carries a bias, as shown by MacRae (1970) and others. Therefore, we are looking, at best, for a range of values, or even an order of magnitude for the information per stimulus. Perhaps, like all magic, the magic of the number seven is only heightened by its elusiveness.

In general, in categorical judgments experiments, the stimuli are presented to the subjects in a variety of ways: single timed stimuli, double pulsed stimuli, etc. The responses are also elicited in a number of ways: as rapidly as the subjects can provide them, taking as long as needed, etc. It should be remembered that most categorical judgment experiments described in this paper conform to one protocol: stimuli are applied and held constant for a "long" period of time. In actual practice, adaptation is 90% complete in less than 10 sec, and sometimes less than

2 sec. So "long" implies "approaching adaptation time." Therefore, we are not considering here the type of experiment described by Laming (1968) and others as a "choice-reaction" experiment. Such experiments are, of necessity, more difficult to analyze. It is not surprising that he found difficulty in applying the Shannon-Wiener theory of information of these experiments. We must stop short, however, of agreeing with Laming's statement (1973) that "the notions of Shannon's theory are entirely inappropriate to human behaviour."

The theory of perception which is being pursued in this paper was not formulated specifically for the purpose of understanding categorical judgments. It was conceived as a means of satisfying certain philosophical requirements (Norwich, 1977, 1978). The new theory differs substantially from other theories of perception, in that it does not, in the literal sense, regard the sensory neurons as providing "information" about the external world to the brain. The very fact that a sensory neuron is transmitting at a rate above its noise level provides the brain with information that heat, light, etc., is impinging on the organism. But the *rate* of impulse transmission relays to the brain primarily the uncertainty residing at the sensory receptor concerning some aspect of the stimulus. Therefore, for long periods of observation of a steady stimulus (τ large in Equation 15), the impulse transmission rate, F , approaches zero (above its spontaneous transmission rate). Stated another way: an organism cannot perceive a certainty. The underlying philosophy states that perception is always relative to the expectation of the organism.

Although Equation 15 states that F must approach zero for large values of τ , it is clear from many experimental observations that the impulse transmission rate does not always fall to the background level in an adaptation curve. There is, in fact, a limitation to the local receptor "memory"—the limited capacity of any sensory unit to store the outcomes of recent events. Thus, there is a maximum value of time, τ , which can enter Equation 15, etc. Another way of expressing this limitation (Norwich, Note 1) is the progressive decrease in the ability of the sensory unit to detect the passage of time by recording the occurrence of external events. When the adaptation curve has reached a plateau and is no longer declining, one instant of time is indistinguishable from another to the sensory unit. Relative to this sensory unit, time is stationary.

Summary and Overview

The acquisition of sensory information by the perceiving system is never instantaneous. It is a process that is progressive with time. As information is acquired, uncertainty about the nature of the stimulus falls. Information as a reduction in uncertainty is a concept elaborated by Shannon. The uncertainty

here has been called the *entropy of equivocation* and has been represented by the symbol H_s .

The process of progressive acquisition of information was classified as *psychophysical* or *physiological* in nature. The "psychophysical" cases were nonetheless physiological, but because they dealt with more complex perceptual phenomena, were not readily dealt with on a neuronal basis. It was shown how the perception of nearly any quantity (hue, pitch, point separation, etc.) that may be associated with a categorical judgment process, could be associated as well with a psychophysical adaptation curve, which is just a graph of entropy, H_s , against time, τ . The exact shape of this curve could be calculated using the techniques of Garner and Hake (1951). The maximum transmitted information, $I_{t \max}$, was simply the difference between the maximum and minimum values of H_s in the psychophysical adaptation curve.

The process of progressive acquisition of sensory information was then examined from a purely neurophysiological point of view. A steady sensory stimulus was applied to an isolated sensory receptor, and the firing rate, F , of the primary afferent neuron described an adaptation curve. The physiological adaptation curve of F against time possessed the same general properties as the psychophysical adaptation curve of H_s against time.

We then introduced the hypothesis that firing rate, F , was proportional to entropy, H_s : $F = kH_s$ (Equation 6). This hypothesis brought the two types of adaptation curve into apposition. For example, it does not now matter whether a spectrum of intensities of light is broken into arbitrary categories for a categorical judgment experiment or whether the light is shined upon an isolated sensory receptor. In each case, the maximum stimulus uncertainty is governed by the photon statistics of the light signal. If the light sensors of the subject are the limiting factor in determining the total transmitted light information, both psychophysical and physiological experiments should give the same value for $I_{t \max}$.

We can now see that both types of adaptation curve are, in essence, the same. The psychophysical adaptation curves are mediated physiologically by neuronal activity, and the individual neurons involved are limited in their capacity to carry information. It is, perhaps, not surprising that the "magical" $I_{t \max}$ has about the same value for both types of curve.

The advantage in using the neurophysiological adaptation curves is that they were obtained independently of any direct measure of transmitted information. Since, from the general theory of information we were able to obtain the mathematical form of the function $F = kH_s$, we could curve-fit these physiological adaptation curves to the function F and thus obtain a value for the proportionality constant, k .

The maximum information transmitted during the sensory process, $H_{s \max} - H_{s \min}$, was then simply equal to $(F_{\max} - F_{\min})/k$. From Tables 1 and 2 it can be seen that for neurons of the frog and cat, the transmitted information per stimulus lies in the range of 2.5 to 4 bits. One is led to suspect, therefore, that if a frog or cat could participate in an experiment involving absolute judgments of the magnitude of a force or the intensity of a tone, we would observe that the transmitted information per stimulus does not exceed this range of 2.5 to 4 bits, which is the quantity relayed to the central nervous system. In this regard, a remark of E. D. Adrian should be recalled (Granit, 1955). "Comparing," said Adrian, "the impulse discharge in an eel's optic nerve and the brightness of a visual image in man may be like the comparison of chalk with cheese, but it seems to be justified by the likeness of the two sets of curves."

Before proceeding with the analysis of Miller's magical number, some further development of the H_s function was required. It was cast into two simpler forms, one for higher and one for lower signal-to-noise variance ratios. The higher ratio gave rise to the (logarithmic) Weber-Fechner law; the lower ratio gave rise to Stevens' power law. In each case we saw that three important quantities were all measured by the same mathematical function: the firing rate of a neuron, the intensity of sensation (sensory effect), and the entropy of equivocation (uncertainty) of a neuron or a human subject. All three quantities were measured by H_s .

Using the power function approximation, the equal magnitude law of Bunsen and Roscoe was obtained by a straightforward process.

Using the logarithmic approximation, it was shown that Miller's magical amount of information was derivable from the ratio τ_2/τ_1 ; the time of, say, 90% adaptation, to the time of maximum response.

Finally, we were able to account for the ability of an observer to discriminate between very small differences in stimulus intensity, in contradistinction to his limited ability to make absolute judgments. The ability to discriminate between two stimuli of nearly the same magnitude is related to the smallest perceptible amount of information transmitted. A neuron either fires or it does not. The effect is quantized: all or none. Using $F = kH_s$, we were able to translate "one impulse" into "one difference limen of transmitted information." In this way we could estimate the rather large number of distinctions that can be made over the audible range of intensities or tone. As a corollary, the magnitude of the constant, k , for any sensory modality, could be related quantitatively to our ability to discriminate.

The entropy function, H_s , has relevance in many areas of psychophysics and sensory physiology. It is a function relating sensory magnitude and neuronal

activity to both stimulus magnitude and stimulus duration. This function, therefore, predicts the outcome of many experiments that involve these three variables. Only the first few milliseconds of the experiment are exempt from control of the H_s function, since in this time interval, the distribution of means is not necessarily normal. This function unifies many laws of psychophysics and accounts, in large measure, for the magical number of G. A. Miller.

REFERENCE NOTE

1. Norwich, K. H. *On the relationship between physics and physiology* (Report No. 18, Institute of Biomedical Engineering). Toronto: University of Toronto, 1978.

REFERENCES

- CUDDY, L. Training the absolute identification of pitch. *Perception & Psychophysics*, 1970, 8, 265-269.
- DURLACH, N. I., & BRAIDA, L. D. Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 1969, 46, 372-383.
- FULGOSI, A., & BARTOLOVIĆ, B. Practice and some methodological problems in unidimensional information transmission. *Perception & Psychophysics*, 1971, 10, 283-285.
- GALAMBOS, R., & DAVIS, H., The response of single auditory-nerve fibers to acoustic stimulation. *Journal of Neurophysiology*, 1943, 6, 39-57.
- GARNER, W. R., & HAKE, H. W. The amount of information in absolute judgments. *Psychological Review*, 1951, 58, 445-459.
- GRANIT, R. *Receptors and sensory perception*. New Haven, Conn: Yale University Press, 1955.
- HARPER, R. *Human senses in action*. Edinburgh: Churchill Livingstone, 1972.
- HARTLINE, H. K., & GRAHAM, C. H. Nerve impulses from single receptors in the eye. *Journal of Cellular and Comparative Physiology*, 1932, 1, 277-295.
- LAMING, D. R. J. *Information theory of choice-reaction times*. London: Academic Press, 1968.
- LAMING, D. R. J. *Mathematical psychology*. London: Academic Press, 1973.
- LUCE, R. D., GREEN, D. M., & WEBER, D. L. Attention bands in absolute identification. *Perception & Psychophysics*, 1976, 20, 49-54.
- MACRAE, A. W. Channel capacity in absolute judgment tasks: An artifact of information bias? *Psychological Bulletin*, 1970, 73, 112-121.
- MATTHEWS, B. H. C. The response of a single end organ. *Journal of Physiology*, 1931, 71, 64-110.
- MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, 63, 81-97.
- MOUNTCASTLE, V. B. *Medical physiology* (14th ed.). St. Louis: Mosby, 1980.
- NORWICH, K. H. On the information received by sensory receptors. *Bulletin of Mathematical Biology*, 1977, 39, 453-461.
- NORWICH, K. H. An hypothesis on information, memory and perception. *Medical Hypotheses*, 1978, 4, 156-164.
- NORWICH, K. H. The information content of a steady sensory stimulus. *Digest of the Combined Meetings: XII International Conference on Medical and Biological Engineering. V International Conference on Medical Physics*. Part 1 19.2, 1979.
- SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*, 1948, 27, 623-656.
- SIEGEL, W. Memory effect in the method of absolute judgment. *Journal of Experimental Psychology*, 1972, 94, 121-131.

NOTE

1. We observe that this ratio could also have been written in the form

$$\frac{\int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / \frac{1}{2}(\tau_{\text{eq}} - \tau_{\min})}{\Delta_{\mu} \int_{\tau_{\min}}^{\tau_{\text{eq}}} H_s d\tau / \frac{1}{2}(\tau_{\text{eq}} - \tau_{\min})}$$

Since the shape of the H_s (adaptation) curve is roughly triangular, this ratio is approximately equal to

$$\frac{\text{altitude of triangle}}{\Delta_{\mu} (\text{altitude of triangle})} = \frac{\text{total sensory effect}}{\text{change in sensory effect}} \quad (44a)$$

APPENDIX A

Fitting Physiological Adaptation Curves to the Entropy Function, H_s

Adaptation curves of the type shown in Figure 2 and given numerically in Tables 1 and 2 can be curve-fitted to the entropy function, as given by Equation 15:

$$F = \frac{k}{\log_e 4} \log_e \left(1 + \frac{B}{\tau} \right). \quad (15)$$

F is the firing rate of the sensory neuron, and τ is time. B and k are two constants to be determined by the curve-fitting procedure. If one has available a good computer program for fitting nonlinear functions, no problem exists. Alternatively, the following procedure is very simple and can be executed on a programmable desk calculator.

Equation 15 can be rearranged to give

$$\exp [(F \log_e 4)/k] = 1 + \frac{B}{\tau}. \quad (A1)$$

Setting

$$y = \exp [(F \log_e 4)/k] \quad (A2)$$

and

$$x = 1/\tau, \quad (A3)$$

we obtain

$$y = 1 + Bx. \quad (A4)$$

Suppose that we assume a value for the constant, k . Nearly any real, positive number will do. Then x and y can be calculated from the experimental data, and the theory, as embodied by Equation A4, predicts that x and y will define a straight line whose y intercept is 1 and whose slope is B . The value of B which best conforms the data to a straight line by the least squares criterion is given by

$$B = \frac{\sum_i x_i y_i - \sum_i x_i}{\sum_i x_i^2}. \quad (A5)$$

Since the initial selection of a value for k has been arbitrary, x and y will not likely lie on a good straight line, and the sum of squares of errors will be large. We then change the value of k , either increasing it or decreasing it, and reevaluate the sum of squares of errors. If this sum is smaller than before, we continue to change the value of k in the same direction; if the sum is larger than before, we change k in the opposite sense. This searching procedure is continued until a minimum sum of squares of errors is obtained. In this way, a "best" pair of values for B and k is found. The final value for y , when plotted against x , will probably give quite an acceptable straight line. The magical number is then estimated from the equation

$$\text{number of categories} = 2^{(F_{\max} - F_{\min})/k}.$$

I have encountered only one problem in this curve-fitting process. When analyzing adaptation curves of low-intensity stimuli, B is found to be small (as may be seen from Equations 13 and 16), so that Equation 15 is approximated using the Taylor expansion of $\log_e(1 + B/\tau)$ in the first power:

$$F = kB/(\tau \log_e 4). \quad (\text{A6})$$

F and τ define a rectangular hyperbola, and, in this case, only the product kB can be obtained from the adaptation curve.

I have carried out much of the curve-fitting in this way with a Hewlett Packard 9825A desk calculator. Some of the data have been fitted using Powell's algorithm with an IBM system 370 computer.

APPENDIX B

Information Theory: Derivation of Equation 9

Let $p_{SN}(I)$ be a continuous probability density function for stimulus intensity together with noise. That is, $p_{SN}(I)dI$ is the probability that signal with noise will have a value between I and $I + dI$. This function is normalized so that

$$\int_{-\infty}^{\infty} p_{SN}(I) dI = 1. \quad (\text{B1})$$

Negative values of I do not exist. For reasons discussed in the main text, we are interested in the case in which p_{SN} is a normal distribution, and particularly in the cases in which

$$p_{SN} = 0, \text{ very nearly, when } I < 0. \quad (\text{B2})$$

Let $p_N(I)$ be a probability density function for noise, which is also normal or Gaussian, and is similarly normalized.

Sampling the stimulus signal will provide a reduction in the uncertainty or entropy of the stimulus intensity, I , and this reduction in entropy is equal to the gain in information that was produced by the sampling process. The entropy, H_s , is given by the equation

$$\begin{aligned} H_s \text{ natural units} &= - \int_{-\infty}^{\infty} p_{SN}(I) \log_e p_{SN}(I) dI \\ &+ \int_{-\infty}^{\infty} p_N(I) \log_e p_N(I) dI. \end{aligned} \quad (\text{B3})$$

This equation is based on a mathematical representation of thermodynamic entropy which issued from the work of L. Boltzmann, was generalized from physics to communications theory by Shannon (1948), and is now being extended into psychophysics.

We see from Equation B3 that if the pure stimulus were absent, so that p_{SN} was equal to the noise distribution p_N , H_s would then be equal to zero.

Let us now evaluate the integrals in Equation B3. Suppose we take any normal or Gaussian function, $p(I)$:

$$p(I) = \frac{1}{\sqrt{2\pi s^2}} e^{-(I-\mu)^2/2s^2}. \quad (\text{B4})$$

s^2 is the variance of the distribution; μ is the mean. Then we have

$$\begin{aligned} &\int_{-\infty}^{\infty} p(I) \log_e p(I) dI \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-(I-\mu)^2/2s^2} \log_e \frac{1}{\sqrt{2\pi s^2}} e^{-(I-\mu)^2/2s^2} dI \\ &= \log_e \frac{1}{\sqrt{2\pi s^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-(I-\mu)^2/2s^2} dI \\ &\quad - \frac{1}{2s^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-(I-\mu)^2/2s^2} (I-\mu)^2 dI \end{aligned} \quad (\text{B5})$$

The value of the first integral is unity because $p(I)$ is normalized, as shown in Equation B1. The second integral represents the expectation of $(I-\mu)^2$, and is equal to the variance s^2 . Hence,

$$\begin{aligned} - \int_{-\infty}^{\infty} p(I) \log_e p(I) dI &= -\log_e \frac{1}{\sqrt{2\pi s^2}} + \frac{1}{2} \\ &= \frac{1}{2} \log_e (2\pi s^2) + \frac{1}{2} \\ &= \frac{1}{2} \log_e (2\pi e s^2). \end{aligned} \quad (\text{B6})$$

Using Equation B6, we can evaluate the entropy function, H_s , as given by Equation B3. Since the variance of p_{SN} , the probability density function for signal plus noise, is $\sigma^2 + N^2$, and the variance of p_N , the probability density function for noise, is N^2 , we have, using Equations B3 and B6,

$$\begin{aligned} H_s \text{ natural units} &= \frac{1}{2} \log_e [2\pi e(\sigma^2 + N^2)] - \frac{1}{2} \log_e (2\pi e N^2) \\ &= \frac{1}{2} \log_e \frac{\sigma^2 + N^2}{N^2} = \frac{1}{2} \log_e \left(1 + \frac{\sigma^2}{N^2} \right). \end{aligned} \quad (\text{B7})$$

If we replace σ^2 by σ^2/n , the variance of a distribution of means of samples of size n , Equation B7 gives us Equation 9. In so doing, we deviate from the canon of information theory as set forth by Shannon, to formulate a variation of the entropy function potentially useful in psychophysics.

Equation 9 is designed to give the information transmitted to a sensory receptor which is receiving a signal known to be of constant intensity, in the presence of random, Gaussian noise. Only time and testing will determine the utility of this function.

GLOSSARY OF SYMBOLS

Symbol	Meaning	Introduced in Equation			
b	proportionality constant	16	p_{SN}, p_N	probability density functions for signal with noise and for noise alone	8
B	a constant	13	Q	number of categories	
e	2.718 . . . the base of natural logarithms		s	as a subscript indicates stimulus as a variable indicates the variance of a normal distribution in Appendix B	B4
F	impulses per unit time in a neuron	6	x, y, z	functions defined for convenience, and each used on one occasion only	
H_s	stimulus equivocation <i>or</i> receptor entropy or uncertainty	1	γ	proportionality constant	11
i	summation index	A5	Δ_μ	operator giving the small change in an expression associated with a small change in μ	43
I	stimulus intensity		μ	mean value of signal intensity	16
I_s, I_t	stimulus information, transmitted information	1	σ^2	variance of a signal (stimulus)	9
k	proportionality constant	6	τ	the time that has elapsed since the instant of application of the stimulus	
m	an exponent	16	$\tau_{\min}$	the time to the first impulse in response to a stimulus	42
max, min	subscripts denoting maximum and minimum values of a variable		τ_{eq}	the time to equilibrium impulse rate following a stimulus	42
n	number of samples of the signal taken by the receptor	11	τ_1	same as $\tau_{\min}$	29
N^2	variance of noise signal	9	τ_2, τ_3	each exceeds τ_1 by the duration of a flash	29, 33
p	represents a general normal or Gaussian function				

(Received for publication May 3, 1980; revision accepted February 19, 1981.)