

CHAPTER 14

ODDS AND n 's, AND THE MAGICAL NUMBER $\log(2\pi)$ BITS OF INFORMATION

RETROSPECTIVE: THE GENEALOGY OF $F = kH$

With this chapter, we come to an end of the “Validation” block in the flow diagram of Figure 1.2. It is, perhaps, useful to take inventory of what has been accomplished.

Equation (2.6), $F = kH$, has been put forward by way of a postulate, relating the perceptual variable, F , to the entropy, H . Using a simple model of the stimulus signal, we derived

$$H = H(\beta, n; I, t) = \frac{1}{2} \ln(1 + \beta I^n / t) . \quad (9.19)$$

Combining Equations (2.6) and (9.19) gave us

$$F = F(k, \beta, n; I, t) = \frac{1}{2} k \ln(1 + \beta I^n / t) . \quad (9.20)$$

In Chapters 10 through 13, we subjected Equations (9.19) and (9.20) to a series of mathematical operations, and, bearing in mind the interpretation of H as an information, derived a host of laws of sensory science that had earlier been observed experimentally purely as empirical rules. We also derived theoretically a number of laws that have not been observed experimentally but might well be in the future. Schematically, much of these chapters might be summarized as in the “Genealogy” charts that follows (p. 165-6).

F JUST ABOVE THRESHOLD FOR DETECTION: DISTINCTION BETWEEN $F(\text{PSYCHOPHYSICAL})$ AND $F(\text{NEURAL})$

We continue here our discussion of “ F at the threshold,” begun in Chapter 9. The equation $F = kH$ must be scrutinized at small values of H . The theory, as developed in Chapter 13, requires that a minimum quantity of information, ΔH , be received by the receptor before the perceiver (psychophysically) reaches the threshold of detection. Clearly, then, for $H < \Delta H$ at the receptor, we must have $F(\text{psychophysical}) = 0$. We seem to be led to the position:

$$F(\text{neuronal}) = k' H(\beta; I, t) , \quad (2.6) / (14.1)$$

but

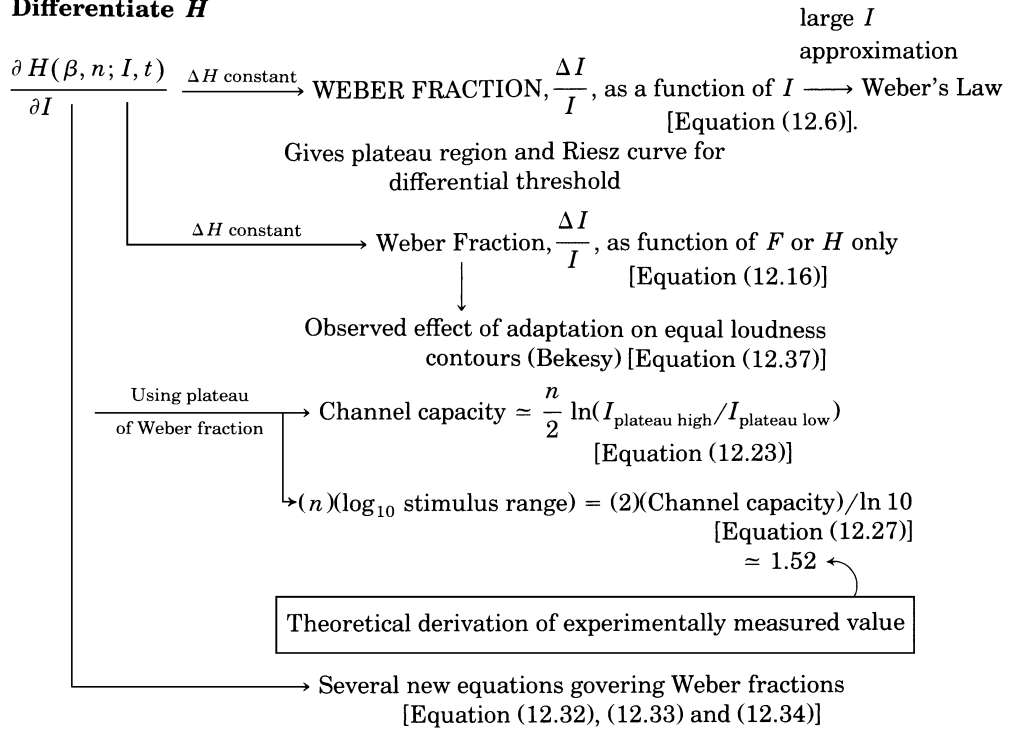
$$\begin{aligned} F(\text{psychophysical}) &= k H(\beta / \xi; I, t / \xi), & \text{if } H(\beta; I, t) \geq \Delta H , \\ &= 0 & \text{otherwise,} \end{aligned} \quad (14.2)$$

where ξ is the time dilatation factor introduced in Equation (13.28).¹ Equation (14.2) is written in contradistinction to

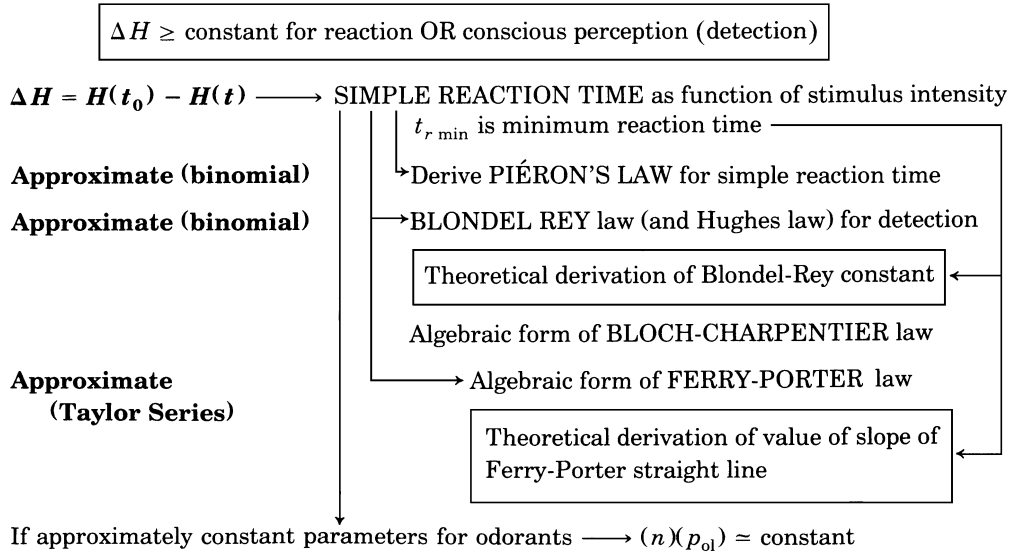
$$F(\text{psychophysical}) = k[H(I - I_0, t)] , \quad (14.3)$$

THE GENEALOGY OF $F = kH$ (Continued)

Differentiate H



$H(\beta, n; I, t)$ used purely algebraically



In Figure 14.1c, we have a semi-quantitative explanation for the well-known Broca-Sulzer effect: at high light intensity (luminance) levels, the momentary brightness shortly after a flash of brief duration (0.03 to less than 1.0 s) appears greater (as much as five-fold) than the brightness of a maintained light stimulus (1 second or more) of the same intensity as the flash. The curve of Figure 14.1c is of the same form as those in Figures 5 and 6 of Broca and Sulzer's original paper (1902), which we can now interpret as adaptation curves. In fact, applying Equation (11.10) for adaptation data to the Broca-Sulzer curves of greater amplitude (far from threshold) provides an estimate of information transmitted per stimulus. For example, for Figure 5 we have, approximately,

$$\text{information} = \log_2 \sqrt{0.5/0.03} = 2 \text{ bits/stimulus.}$$

Again, we have used duration of stimulus as an approximation of time since onset of stimulus, the variable used in the entropy equation.

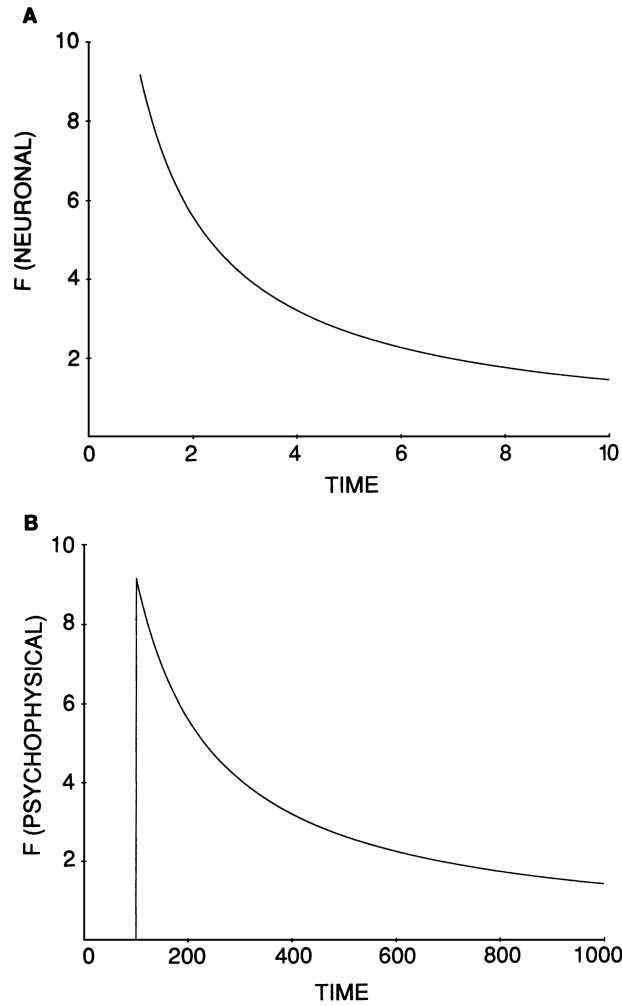


Figure 14.1 a&b (a) Schematic of a neuronal adaptation curve. The curve was generated by Equation (11.3), $F = kH = \frac{1}{2}k \ln(1 + \beta I'^n/t)$, with $k = 10$ and $\beta I'^n = 1.5$. t_o , the earliest neuronal response, occurs at 1 unit of time. ΔH , the total information transmitted by this adaptation process, can be determined using the methods discussed in Chapter 11. The information transmitted by the portion of the curve illustrated is, then, equal to $-\frac{1}{2} \ln[(1 + 1.5/10)/(1 + 1.5/1)] = 0.388$ natural units.

(b) Psychophysical adaptation curve derived from the neuronal curve of Figure 14.1a by use of a “time dilatation” factor, $\xi = 0.01$ (see Box 13.1). That is, using Equations (B.2a) and (B.2b) from Box 13.1, we write $\beta(\text{psychophysical}) = (1/\xi)\beta(\text{neuronal}) = 100\beta(\text{neuronal})$, and $t_o(\text{psychophysical}) = 100t_o(\text{neuronal})$. Therefore, this curve was generated by Equation (11.3) with $k = 10$ and $\beta I'^n = 1500$, beginning with $t_o = 100$ time units. We note that the psychophysical adaptation curve is identical in shape to the neuronal curve, but shifted forward in time, and spread out in time. We note also that the psychophysical curve does not begin until information, ΔH , has been transmitted by the neuronal curve. That is, $t_o(\text{psychophysical}) > 10$ time units. The information transmitted by the psychophysical process is equal to

$$-\frac{1}{2} \ln\{[1 + (1.5)(100)/1000] / [1 + (1.5)(100)/100]\} = 0.388$$

natural units, as before.

Figure 14.1a and b are in keeping with the experimental findings: neuronal adaptation processes occur rapidly following stimulus onset, while psychophysical processes (conscious awareness of the stimulus) are delayed in time and dilated in time (cf. Chapter 13). Both types of adaptation curve, when analyzed by the methods of Chapter 11, yield the same value for transmitted information.

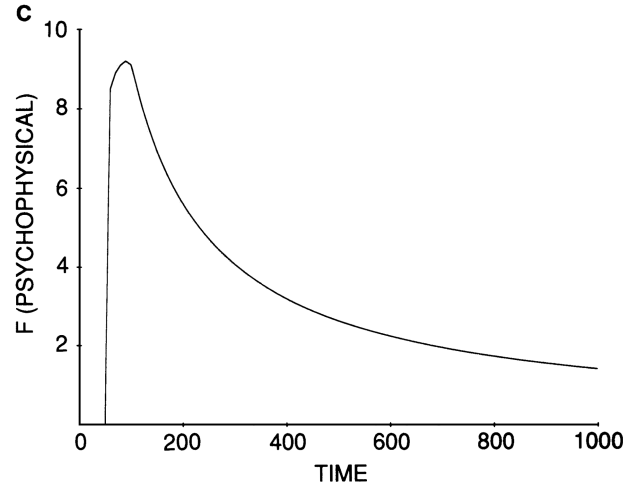


Figure 14.1c Schematic. Recalling that no conscious awareness of the stimulus begins until the neuronal adaptation process has transmitted its capacity of ΔH units of information, and allowing that the psychophysical adaptation process relies on inputs from a number of neurons, there may be a range of thresholds, $\Delta H \pm \epsilon$, which would have the effect of blunting the curve of Figure 14.1b to give this effect. Curves of this type have been observed experimentally.

Equation (14.2) is still, in itself, not adequate because it presupposes that when the receptor finally receives information, ΔH , the psychophysical response “jumps” suddenly from zero (imperceptible stimulus) to $kH(I, t/\xi)$, which may be substantially greater than zero. Such jumps at threshold are not in keeping with experience. We require, therefore, something like¹

$$F(\text{psychophysical}) = \begin{cases} k[H(\beta/\xi; I, t/\xi) - \Delta H], & \text{if } H(\beta/\xi; I, t/\xi) \geq \Delta H \\ = 0 & \text{otherwise,} \end{cases} \quad (14.4)$$

in keeping with Equation (9.22) and with the findings of Lochner and Burger (1961). This equation looks complicated because we are using neuronal β - and t -values with the time-expansion factor, ξ ; but the equation just states that

$$F = \begin{cases} k(H - \text{threshold}), & \text{if } H \geq \text{threshold} \\ = 0 & \text{otherwise.} \end{cases} \quad (14.4a)$$

This equation is struck by way of conjecture, because it eliminates the “jump start” of sensation and it seems to encompass much that we have learned about neuronal / psychophysical information receipt. There are, however, insufficient data available to test and refine the equation. So I leave it with the reader as the state of the art in the autumn of 1992.

EXPANDING THE HORIZONS BEYOND “INTENSITY” AND “DURATION”

The sole independent variables used in the current theoretical development are intensity of the stimulus (regarded as constant) and time since onset of the stimulus. The cardinal dependent variable, the entropy, H , has been regarded as a function of these two variables only. While I think most readers would agree that the initial mathematical development has been complicated enough with just these two independent variables, there would also be tacit agreement that the experience of sensation and perception is considerably broader than single, monochromatic step functions.

There are many directions in which the entropic view of perception might be extended. We have seen in the previous chapter, for example, how simple reaction time is affected not only by the entropy of stimulus intensity, but also by the entropy of stimulus onset time (Klemmer’s “foreperiod

variability"). We have also seen that the entropy of the complexity of a task (Hick) affects reaction time. So the entropy of stimulus intensity is just one dimension of a multidimensional phenomenon.

We have regarded the "law of sensation," $F = F(k, \beta, n; I)$ [t constant], as a function involving only three fixed parameters, k , β , and n , which is an egregious oversimplification. Even if we confine our attention to single, constant stimuli of physical magnitude, I , disregarding masking stimuli and the like, three fixed parameters will not suffice. Marks and Warner have shown brilliantly (1991) how *context* affects F , not just marginally, but emphatically, where context refers to the ensemble or set of possible stimulus values. Lockhead (1992) emphasized the importance of "assimilation" [subject's values of F (psychophysical) tend to reflect the value of F in the trial just preceding]. So the constrained dependency of F on just 3 parameters and 2 independent variables constitutes only a simplified beginning. I believe that the $F = kH$ concept will mature only when investigators carry it with them into the laboratory, and design experiments to test and expand it.

In Chapter 17, we shall explore various other extensions and applications of the entropy concept of perception.

THE EXPONENT n

I must confess to a certain antipathy² toward this exponent, n . I believe I have harbored this feeling of ill will ever since I found it necessary to introduce the exponent by means of the empirical equation (9.14). After all, one of the reasons for developing the H -function was to be able to replace a host of empirical equations with equations that were logically derived. And here, at the very base of the logical derivation, is just the sort of equation I was trying to get rid of.

"But no matter," I reasoned slyly. "I'll just hide the unsavory little n , multiplied and divided by a hoard of other parameters (you've heard of the 'lumped parameter model'; everybody uses them) in a new, synthesized parameter called ' θ ' or ' Λ^2 ' (Greek letters always lend an air of authority to your parameters), and no one will ever see the n again!" But, as the fates would have it, the little demon, n , kept escaping from its parametric prison. No sooner would I bury it in a compound parameter³ ($kn = a$, Equation (10.5, 10.5a)), then it would resurrect itself to stand alone in the Weber fraction [Equation (12.6)], and announce its primacy on the plateau of the Weber fraction in Equation (12.27):

$$n \cdot \ln(\text{range of plateau}) \simeq 2 \delta H \simeq 3.5 \text{ natural units,} \quad (14.5)$$

or

$$\text{range of plateau} \simeq e^{3.5/n}. \quad (14.6)$$

I would hide its light under a bushel in Equation (13.6), making I_{\min} the primary variable, only to find that the n had immolated its bushel to appear again as a stand-alone parameter in the equation for simple reaction time (Equation (13.7)), and in Piéron's equation (13.14).

Would no one rid me of this troublesome power? I laid awake nights plotting, but my graphs came to naught.

If Equation (9.14) had been born of the marriage of logical ancestors, I would have no objection to the numerousness of its progeny. But the birth of n in the illegitimate cauldron of empirical iniquity rankled me to my puritanical roots, as I suffered again its brazen countenance in the form of a fundamental variable in the Blondel-Rey constant (Equation (13.42)).

How far could n drift from its humble origins in the power law of sensation? At least as far as the slope of the straight line in the Ferry-Porter law (Equation (13.51)). And, of course, the approximate law of olfactory thresholds would have to be

$$n \cdot p_{\text{ol}} = k. \quad (13.55)$$

Well, if we cannot displace n , then more power to it! Let's explore various appearances of the exponent n to see if its numerical values are consistent with each other. In particular, let's explore the values of n measured for the sense of hearing.

THE EXPONENT n AT VARIOUS AUDITORY FREQUENCIES

We have learned that there are values of the exponent, n , that are characteristic of each of the modalities of sensation. The measured value of n is, moreover, modulated within a modality by numerous factors such as the mode of measurement [magnitude estimation vs. magnitude production (*e.g.* Meiselman *et al.*, 1972)], and the duration of the stimulus (*e.g.* Ueno, 1976, shows that n for vision falls by a factor of about 40% for brief stimuli). The value of n for intensity of pure tones also changes with the frequency of the tone tested. In classical psychophysics, n can be measured only by carrying out an experiment relating subjective magnitude to physical magnitude (law of sensation). However, because of the appearance of the exponent, n , in many of the equations of the entropy theory, we are afforded the means of estimating the value of n by various techniques that do not involve subjective magnitudes. One of the potential applications of the entropy theory of perception is this objective (or more objective) means it provides for the evaluation of the power function exponent. Comparing the psychophysical with the entropic evaluation of n for audition provides a further experimental test of the validity of the entropy theory.

Let us examine three ways by which n can be measured at different auditory frequencies.

(i) *Measuring n for pure tones psychophysically for many frequencies*

One could, of course, proceed as follows. At 1000 Hz, measure subjective magnitude, F , corresponding to many different physical magnitudes (sound intensities), I ; plot $\ln F$ vs. $\ln I$, and calculate $n(1000)$. Then repeat the process at 2000 Hz to obtain $n(2000)$, etc. Equivalently, one can take advantage of charts and tables of equal loudness contours, such as those provided by Robinson and Dadson (1956), a technique that was used by Lochner and Burger (1962). In this method, one measures F vs. I only once: for 1000 Hz, giving $n(1000)$. Represent these measured F -values as $F(I_{1000})$ (see *sone* vs. *phon*, Evans, 1982). For example, when $I_{1000} = 80$ dB, $F(I_{1000}) = F(80 \text{ dB}) = 15$ sones of loudness (say). Using the equal loudness contours, one can then find for 2000 Hz the intensities, I_{2000} , for which loudness is equal to $F(80 \text{ dB})$. This sounds more complicated than it really is. A glance at Table 14.1 should clarify the issue. Anyway, in this manner, we can plot a graph of $\ln F(I_{2000})$ vs. $\ln I_{2000}$, whose slope is equal to $n(2000)$. In similar fashion, we can obtain values of n for all frequencies in the physiological range of hearing, using only one set of measurements of loudness. A graph of n vs. tone frequency measured using equal loudness contours is given in Figure 14.2. The slopes (values of n) were found by curve-fitting using a simplex program.

(ii) *Measuring n for pure tones using the Weber fraction*

The Weber fractions measured by Riesz (1928) for audition are well described by Equation (12.6), which gives values of $\Delta I/I$ as a function of n . In fact, Equation (2) of Riesz, which he used empirically to describe his data, is identical with our Equation (12.6):

$$\Delta I/I = (2\Delta H/n) + (2\Delta H/n\gamma)(1/I)^n. \quad (12.6) / (14.7)$$

Table 14.1 Plotting loudness against intensity for different frequencies using equal loudness contours

$I_{1000}(\text{dB})$	$I_{2000}(\text{dB})$	$F(I_{1000})$
10	7	Loudness 1
20	17	Loudness 2
....
120	110	Loudness 12

Note: The third column lists loudnesses measured to correspond with the sound intensities listed in column 1 (for 1000 Hz). The second column lists (for 2000 Hz) sound intensities that are equally as loud as those corresponding in column 1, obtained from a graph or table of equal loudness contours. We could proceed in this fashion to construct a column for every frequency desired in the physiological range. Plotting the loudnesses in the right-hand column against each of the other columns in the table, in turn, provides a set of curves, each defining the law of sensation at a certain frequency.

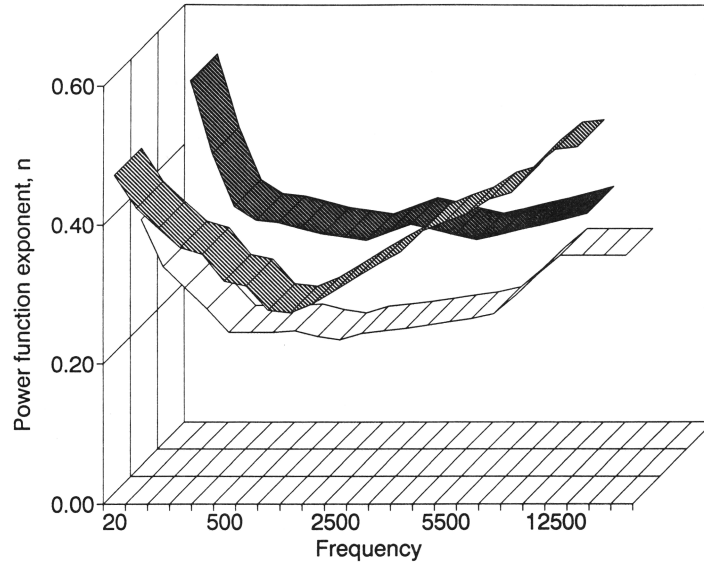


Figure 14.2 The exponent, n , for audition, plotted against frequency of pure tones, by three independent methods.

(i) *Front curve*. Riesz measured differential thresholds (Weber fractions) for a range of auditory frequencies. He fitted his data empirically to an equation that we identify as our Equation (12.6), giving differential thresholds as a function of n . Hence, we can plot n as a function of frequency. (ii) *Middle curve*. Chocholle measured simple reaction times for a range of auditory frequencies. By fitting his data to our Equation (13.7), we can find a value of n for each frequency. (iii) *Rear curve* n is measured psychophysically for pure tones of 1000 Hz. n is calculated for other frequencies using equal loudness contours.

Riesz, using the symbol E for our I , wrote

$$\Delta E / E = S_{\infty} + (S_0 - S_{\infty})(E_0 / E)^r. \quad (14.8)$$

By comparing Equations (14.7) and (14.8) we see that Riesz's r is equivalent to our n . Riesz gave empirical equations for each of S_{∞} , S_0 , and r as functions of frequency, f . We include these empirical equations here for the sake of completeness.

$$S_{\infty} = 0.000015f + 126 / (80f^{0.5} + f), \quad (14.9)$$

$$S_0 = 0.3 + 0.0003f + 193/f^{0.8}, \quad (14.10)$$

and, most importantly,

$$n = r = \frac{244000}{358000f^{0.125} + f^2} + \frac{0.65f}{3500 + f}. \quad (14.11)$$

Since Riesz measured the Weber fraction for a wide range of frequencies, f , we can, using Equation (14.11), obtain values of n at any frequency within this range.⁴ These values of n are plotted against frequency in Figure 14.2.

Just a word on Riesz's place in history. Despite the relatively early date of his work (1928), Riesz's results for a single frequency have definitely been reproduced by later investigators. In particular, the reader is referred to Harris (1963, Experiment III: A Repetition of Riesz), and to the comparison of six groups of investigators shown by McConville *et al.* [1991, Figure 4(b)]. There is, however, some debate about whether $\Delta I / I$ changes as radically with frequency as Riesz's data indicate. Riesz used a method of beats to determine ΔI , while other investigators have used a pulsed tone method (such as method II of Chapter 12). The latter method may have produced the muting of the frequency effect that Riesz had discovered. The pulsed tone results are contrasted clearly with the beat results as shown by Gulick *et al.* [1989, Figure (10.9a)].

I also remind the reader that the mathematical method used to derive the theoretical expression for the Weber fraction in entropy theory, Equation (12.6), utilized a small change in intensity, ΔI , without

introducing any adaptation effect (t was held constant). That is, it was a quasi-static method and as such, it matched Riesz's experimental method rather closely, which probably accounts for the nearly exact matching of the numerical results.

(iii) *Measuring n for pure tones using simple reaction times*

We may see from Equation (13.7), that simple reaction time is expressed in terms of the parameter n . Therefore, by measuring simple reaction time as a function of I for each frequency of interest, 50 Hz, 100 Hz ... we obtain, by curve-fitting, measures of $n(50)$, $n(100)$... Fortunately, Chocholle (1940) published just such data giving simple reaction time *vs.* intensity for many frequencies. Equation (13.7) was curve-fitted to Chocholle's data at each frequency, using a simplex optimization program, and the resulting graph of n *vs.* frequency is shown in Figure 14.2.

From the "ribbon graphs" in Figure 14.2, one can see that the exponent, n , varies in value across the range of auditory frequencies, and the graph of n *vs.* frequency is quite similar when measured directly by psychophysical means, or by means of differential thresholds, or by means of simple reaction times. There are differences between the curves, but I think that the general concurrence of n -values constitutes a very powerful experimental test of the entropy theory. Moreover, as stated above, it opens the door to more objective measurements of n for audition using techniques that do not involve the use of subjective magnitudes.

AN INTENSITY-FREQUENCY FORMULATION OF THE ENTROPY EQUATION FOR AUDITION

Our objective in this section is to try to discover an equation that gives two subjective measures, loudness and pitch, as a function of the independent variables intensity and frequency using the informational approach. To my knowledge, no mathematical relationship has been discovered by anyone, experimentally or theoretically, which achieves this goal. We do not achieve the goal here either. However, I shall present a rather natural reinterpretation of the entropy equation (9.20) that may carry some promise for an avenue of approach.

A steady pure tone can be described completely only when its amplitude, frequency, and phase have been defined. The intensity of the tone is related to its amplitude, and this is the only feature of the tone we have considered hitherto. However, a rather natural way to incorporate frequency into our entropic formulation is to regard t , the time variable, not as the time since onset of the auditory stimulus, but rather as an approximation of *the duration or periodicity of the dominant waveform*. In its simplest interpretation, this periodicity would correspond to that of Schouten's *residue*, but, in principle, it need not be even that restrictive. We shall not stop here to define Schouten's residue theory of pitch perception. It is discussed very clearly by Wightman and Green (1974). Following this line of reasoning, we can redefine " t " as " Δt ," the period of the residue or other dominant waveform. Thus, the entropy equation (9.20) takes on the new form

$$F = \frac{1}{2} k \ln(1 + \beta I^n / \Delta t) . \quad (14.12)$$

Then, if Δt is the reciprocal of the fundamental frequency, ν , we have

$$F = \frac{1}{2} k \ln(1 + \beta I^n \nu) . \quad (14.13)$$

Equation (14.13) probably produces more problems than it solves, but it is not without some saving graces. For ν constant, it reduces to

$$F(\text{loudness}) = \frac{1}{2} k \ln(1 + B_I I^n) , \quad (14.14)$$

where B_I is constant, not changing with I . This equation, then, describes the psychophysical law in the usual manner. It would not describe auditory adaptation phenomena, either monaural or binaural. For I constant, Equation (14.13) becomes

$$F(\text{pitch}) = \frac{1}{2} k \ln(1 + B_\nu \cdot \nu) , \quad (14.15)$$

where B_v is constant, not changing with v . $F(\text{pitch})$, then, is given as a function of frequency, v , to the first power. This function is, in fact, precisely that found for pitch in experiments reported by Lindsay and Norman (1977):

$$\text{pitch} = 2410 \log_{10}(1 + 1.6 \times 10^{-3}v). \quad (14.16)$$

One can interpret v as the frequency of Schouten's residue, but perhaps there is a more general interpretation of v , as demanded by modern theories of pitch (again referring to Wightman and Green, 1974). One must look for a physiological mechanism capable of extracting both intensity and pitch information from the combined equation (14.13). We recall that each equation, such as (14.13), that governs the transmission of information, is of necessity associated with a mechanism for transmission (Chapter 1). I leave these questions with the reader. Perhaps Equation (14.13) is just a curious transmogrification of Equation (9.20).

DE-ADAPTATION[†]

We have analyzed, in some depth, how the process of adaptation to a constant stimulus can be represented within the entropy theory, as a process of acquisition of certainty. However, not a word has been said about the informational process that takes place after cessation of the steady stimulus, resulting in loss of certainty and reversion, eventually, to a status of maximal uncertainty. This latter process might be called "de-adaptation." In studies of vision, it is "dark adaptation." Not a word has been written about this phase of sensation because I do not yet understand it. All attempts I have made to date to encompass the de-adaptation phase into the general theory have failed.

Related to the problem of de-adaptation is the problem of modifying the entropy equation to handle a time-varying stimulus. Some investigators have introduced stimuli in the form, for example, of sinusoids (Bohnenberger, 1981), and have measured the neuronal response. My students and I have not been able to predict the response to such time-varying stimuli from the response to the step input. I should add that it is possible to construct rather complicated and arbitrary criteria whereby a sensory receptor can determine whether a change in mean signal intensity (or variance) is due to random fluctuation in a steady signal (stationary time series), or due to a real change in the mean. The trick is to find a *simple* criterion.

THE "MAGICAL" NUMBER 7: WHENCE AND WHY?

The "channel capacity" of a one-dimensional stimulus, as we have seen, is usually about 2.5 bits of information per stimulus, as determined in an experiment involving categorical judgments. For judgments of simple stimulus intensities, channel capacity seldom exceeds 2.5 bits, and is often less. Since $2^{2.5} = 5.7$, the magical number is probably closer to 6 than 7 for our purposes. One of the most encouraging results of entropy theory has been the extraction of the maximum information per stimulus from simple adaptation data, and the demonstration that this maximum stimulus information is very nearly equal to the corresponding channel capacity for that modality. A rather exciting possibility emerged in Chapter 13, when we were able to demonstrate that the threshold for intensity detection (*not* the threshold for intensity discrimination) was probably equal to the maximum information per stimulus.^{††} That is, the maximum information must be received by the neuron and signaled by $F(\text{neuronal})$ (Figure 14.1a) *before* the stimulus can be detected. Hence $F(\text{psychophysical})$ is displaced forward in time, as shown in Figure 14.1b. So we now do make a distinction between F as measured neurally and F as measured psychophysically, but we do not yet have the whole picture.

However, whence arose the mysterious $\log_2 6 \simeq 2\frac{1}{2}$ bits of information? What is the theoretical meaning of this information capacity – information threshold? It is tempting to define this quantity of information (about 1.75 natural units) from Equation (12.27)/(14.5):

$$\begin{aligned} &\text{Maximum information per stimulus} \\ &= \frac{1}{2}n \cdot \ln(\text{range of plateau of Weber fraction}), \end{aligned} \quad (14.17)$$

where, as always, *range* is equal to *maximum intensity of plateau* divided by *minimum intensity of plateau*. For example, for the sense of taste, $\text{range} \simeq 100$, and $n \simeq 1$, so that *maximum information per stimulus* is approximately equal to $(1/2)(1) \cdot \ln(100) = 2.3$ natural units (which is a little high for taste). However, Equation (14.17) leaves something to be desired. It expresses the maximum information in terms of quantities, such as range and exponent, which are not elementary quantities.

Is maximum information per stimulus characteristic of the human species, or is it a limit imposed on all perceiving organisms? In a recent study by Norwich and McConville (1991), maximum stimulus information was calculated for insects and arachnids from adaptation data, and the values obtained were similar to the values found in the human species. If we encountered an intergalactic visitor,^{†††} would its perception be limited to the same quantity of information per stimulus?

I have not been able to produce a perfectly general derivation of the magical number 6, despite the many years I have pondered it. There is, however, an interesting, although restricted, derivation that I would like to introduce.

I ask that you let your mind be plastic, and stretch your thinking to encompass the idea of a hypothetical visual “perceptor,” an abstract entity capable of perceiving light stimuli of low intensities. I am avoiding the term visual “receptor” because that term would admit a photoreceptor of the type found in a laboratory photometer or light meter, and I am interested in a unit which can perceive; that is, which can make a selection from among alternatives (Chapter 2) and signal its uncertainty in accordance with an entropy function. In particular, I ask that you consider a very sensitive visual perceptor, one that can detect even a single photon of light.

Let us use Equation (9.10) with m , the number of samplings, equal to 1.

$$H = \frac{1}{2} \ln(1 + \sigma_S^2 / \sigma_R^2), \quad (14.18)$$

where σ_S^2 and σ_R^2 are the variances of the external light signal and the internal reference signal respectively.⁵ Since the distribution of photons in illumination of low intensity will be governed by the Poisson distribution; and the smallest non-zero mean is one photon; and since for the Poisson distribution, variance is equal to the mean; therefore variance is equal to one photon. If we approximate the Poisson distribution for discrete variables by a probability density function for continuously distributed variables, the variance must then take on the dimensions of the square of the random variable.⁶ Let the continuous random variable be light energy, which occurs in bundles of energy, $h\nu$, by the quantum theory, where h is Planck's constant and ν is the frequency of the light wave. Thus

$$\sigma_S^2 \simeq (h\nu)^2 \quad (14.19)$$

corresponds to unit variance. Notwithstanding the fact that the Poisson distribution with unit mean is not well represented by the normal distribution, we shall introduce Equation (14.19) into (14.18) to obtain

$$H = \frac{1}{2} \ln[1 + (h\nu)^2 / \sigma_R^2]. \quad (14.20)$$

We now ask: What is the maximum value obtainable by H (that is, the maximum stimulus information)? It would seem at first that we can make σ_R^2 indefinitely small, so that H could become indefinitely large. However, on the quantum scale, our liberties are sometimes checked.

Just to be clear, the numerator, $(h\nu)^2$, represents the variance of the signal or stimulus or external photon (light quantum), and the σ_R^2 in the denominator represents the variance of the reference or “noise” or internally generated photon. We recall from Chapters 7 and 9 why the reference signal is held to be necessary when we deal with continuously distributed variables, although one might argue here whether our variables are discrete or continuous.

We invest our visual perceptor with no short-term memory; it can have no “recall” of the arrival of a reference photon in the past. Therefore, the signal photon and the reference photon must arrive at the perceptor simultaneously, since only in the presence of a reference signal can information be measured. This simultaneity will provide us with a restriction on the maximum possible value of H .

Since σ_R^2 , like σ_S^2 , is the square of an energy, we suppose that the reference signal, too, is carried by a single photon, or light quantum. Let us set

$$\sigma_R^2 = (\Delta E)^2. \quad (14.21)$$

This equation, while seemingly a change in symbol, makes a statement of some physical significance. It states that the variance of the distribution of the reference signal, when applied to a single photon, is equal to the square of the “uncertainty” in the magnitude of the energy comprising this photon. Introducing Equation (14.21) into (14.20),

$$H = \frac{1}{2} \ln[1 + (h\nu)^2 / (\Delta E)^2] . \quad (14.22)$$

Now, the photon, or light quantum, is made up of a group of light waves that add together constructively and destructively to produce a *wave packet* that travels through space at the speed of light. Our photon of energy $h\nu$ will be made up by a wave packet which spans a length,

$$\Delta x = c \cdot \Delta t , \quad (14.23)$$

where c is the speed of light and Δt is the time taken by light to travel the distance, Δx . Bohm has shown⁷ that Δx must be at least as great as the wavelength, λ , of the photon, where, from the elementary properties of light waves,

$$\lambda = c/\nu . \quad (14.24)$$

That is

$$c \cdot \Delta t \geq c/\nu , \quad (14.25)$$

or

$$\nu \geq 1/\Delta t , \quad (14.26)$$

so that the frequency of the signal photon is equal to or greater than the reciprocal of the time taken by the wave packet to pass a given point in space. Since we seek the conditions for the near-simultaneous occurrence of two events (arrival of two wave packets: signal and reference), we therefore select the smallest possible value of Δt , corresponding to the largest possible value of $1/\Delta t$, which is equal to ν , by Equation (14.26). Inserting

$$\nu = 1/\Delta t$$

into Equation (14.22),

$$H = \frac{1}{2} \ln \left(1 + \frac{h^2}{(\Delta E \cdot \Delta t)^2} \right) . \quad (14.27)$$

We note that the expression $(\Delta E \cdot \Delta t)$ refers to the reference photon. ΔE is the uncertainty in its energy, and Δt , the time interval of spread of the wave packet of the signal photon, is also the time interval of spread of the wave packet of the reference photon, if the two packets are to arrive nearly simultaneously at the perceptor.

Now, we know from Heisenberg's Uncertainty Principle⁸ that

$$\Delta E \cdot \Delta t \geq h/(2\pi) , \quad (14.28)$$

so that

$$\frac{h^2}{(\Delta E \cdot \Delta t)^2} \leq 4\pi^2 . \quad (14.29)$$

Inserting the latter inequality into Equation (14.27), we have

$$H \leq \frac{1}{2} \ln(1 + 4\pi^2) . \quad (14.30)$$

That is,

$$\begin{aligned} H &\leq 1.850 \text{ natural units} \\ &= 2.67 \text{ bits of information.} \end{aligned} \quad (14.31)$$

The theoretical magical number is then

$$2^{2.67} = 6.36 \text{ categories,} \quad (14.32)$$

corresponding well to the empirical value of about 6.

Since $(1 + 4\pi^2)$ is quite close to $(2\pi)^2$, Equation (14.30) states that H , the maximum information obtainable by perceiving a stimulus of one photon, is given by

$$H \simeq \log_2 2\pi \text{ bits,} \quad (14.33)$$

or, in natural units,

$$e^H = 2\pi. \quad (14.34)$$

What does the above restriction on H mean, physically? Referring to Equation (14.27), it means that if you try to increase the value of H (maximum stimulus information) by decreasing the energy, ΔE , of the reference signal⁹, you will find that Δt , the spread in time of the wave packet carrying the reference photon will, of necessity, increase, by Heisenberg's Principle, thereby frustrating your efforts to increase H . Referring to Figure 8.2 and Equation (14.22), we see that, very roughly, H is the number of times the variance, $(\Delta E)^2$, can "fit into" the variance, $(h\nu)^2$; that is, the number of times the reference variance fits into the signal variance. Any attempt to increase H either by increasing $(h\nu)^2$ or by decreasing $(\Delta E)^2$ will always fail.

How realistic is the above argument? Certainly, the visual "perceptor" is an entirely abstract concept, perhaps bearing no relationship to any real perceptual system. Constraining the argument to one signal photon is restrictive. Approximating the discrete Poisson distribution by the continuous Gaussian distribution is suspect. And, finally the insistence on a memoryless system, where signal and reference photons must arrive within the same time interval ($\simeq 1/\nu \approx 10^{-15}$ s) would seem to remove the argument from the usual time scale of biological events. Nonetheless, the exercise compels us to examine the problem of maximum information carefully, and demonstrates the critical nature of the reference signal. If we regard the light as composed of discrete particles, perhaps we might dispense with the reference photon, since we can calculate information for discrete events without necessarily subtracting a reference signal. But if light is regarded as continuous or wave-like, we must deal with continuous distributions, and the reference signal is required for calculating information. Once in place, the reference signal limits the maximum information transmission. The exercise is also worthwhile because it is as close as we have come to providing a mechanism for an elementary perceptual event. Electrons, scattered by the incoming photons, could be detected. . . But we have not come so far as to suggest a mechanism for generation of the reference signal, so our *perceptual unit* is incomplete.

A BRIEF HISTORY OF THE "COMPLETE" LAW OF SENSATION

We borrow the term "complete" from Nutting (1907), and indirectly from Lehmann (1905) who referred to Fechner's law as "unvollständig," or incomplete.

Our Equation (10.3) was, of course, derived as an equation of entropy. But we have seen that an equation of similar algebraic form has been postulated as a law of sensation by various earlier investigators. Except for Békésy, I can find no evidence that these investigators were aware of each other's work.

In 1873, Delboeuf expanded Fechner's law to obtain Equation (10.20), which is a special case of our Equation (10.3), with $n = 1$.

In 1905, Lehmann wrote a neurophysiological equation of the same form as Delboeuf's.

In 1907, Nutting, adopted Equation (12.11) as an empirical equation representing the Weber fraction for light. Equation (12.11) is of the same form as our entropic Equation (12.6). By integrating, Nutting obtained an equation very similar to our Equation (10.3) for the law of sensation of light. Nutting's process was the inverse of our own: We began with the law of sensation and differentiated to obtain the Weber fraction. Nutting's equation for brightness, given below, is constructed so that $F = 0$ when I lies at the threshold level for detection:

$$F = F_0 \log[1 + P_m (I^n \cdot I_{\text{thresh}}^{-n} - 1)] \quad (14.35)$$

In 1930, Békésy adopted the neurophysiological equation of Delboeuf-Lehmann, Equation (10.3) with $n = 1$ (please refer to Chapter 12).

In 1959, Rushton used the Delboeuf-Lehman-Békésy equation empirically, with the addition of a constant [Equation (10.21)]; see Chapter 10.

In 1977, I derived Equation (10.3), a further generalization of the Delboeuf-Lehmann-Békésy equation, from entropic considerations (please see Chapter 9).

Are there other members of this strange family of *inconnus*?

NOTES

1. Actually, $H(\beta; I, t)_{\text{neural}} = H(\beta/\xi; I, t/\xi)_{\text{psychophysical}}$.

The two H -functions have the same value because both t and β have been divided by the same factor, ξ , in the function on the right-hand side. Recall that in the H -function, we use only the ratio $\beta/t = (\beta/\xi)/(t/\xi)$. This idea is illustrated in Figures 14.1a and 14.1b.

2. Yes, I think that theoreticians can emote toward mathematical entities in much the same manner that normal people express emotions toward other human beings.

3. I particularly objected to its seeking refuge in my initials.

4. The symbol f is used here for frequency, because it is the symbol used by Riesz. However, in the remainder of the chapter, we use the symbol ν .

5. We recognize that the stimulus signal originates externally, or outside the organism, while the reference or “noise” signal originates internally, or within the organism.

6. We let the continuous probability density function be $p(E)$, where E is the energy of incident light. The mean has the dimensions [energy] and the units $h\nu$, and the variance has the dimensions [energy²] and the units $(h\nu)^2$.

7. Bohm (1989) has shown (pp. 92 et seq) that “whenever the light quantum [photon] is observed under conditions in which it is not absorbed, it cannot be localized to a region smaller than λ [that is, one wavelength or time $1/\nu$].” The quantities in parentheses are mine, not Bohm’s. To be rigorous here, we need a model of the perceptor’s mechanism of simultaneous detection of the signal and reference photons. For example, if the photons each scatter an electron, detection of the scattered electrons could not localize the width of the wave packets to a region less than one wavelength. This restriction is important in our argument, since it, in concert with the Uncertainty Principle, ultimately fixes the magnitude of the maximum information transmissible.

8. Bohm (1961, 1989, Chapter 5.11) states that Equation (14.29) may be used only “in any process in which a quantum is transferred from radiation field to matter (or vice versa).” Aharonov and Bohm (1961) caution against the use of Equation (14.29) inappropriately to express “a further uncertainty relation between the *duration* of a measurement and the *energy transfer* to the observed system.” Also, the Uncertainty Principle is sometimes written with a divisor of 4π instead of 2π , which would increase our theoretical value for the magical number by a factor of nearly 2.

9. Actually, $(\Delta E)^2$ is the variance of the reference signal, but by decreasing the variance, you decrease the mean (from the Poisson distribution).

————— 2003 ed. notes:

†. We have, now, made some progress in extending the entropy theory to include the process of de-adaptation and to allow for time-varying stimuli, even sinusoidal in nature. These preliminary results have been published in K. H. Norwich and W. Wong, “A Universal Model of Single-Unit Sensory Receptor Action”, *Mathematical Biosciences*, 125, 83-108, 1995.

††. Well, it might have been exciting and probable, but, alas, it is unlikely to be correct. Later work has suggested that a better guess for this threshold is probably (that word again) $\Delta H = \ln(e/2)$ natural units (refer also to Note ††† in chapter 13). This possibility emerged in the paper entitled “The information of a welcher Weg experiment”, K.H. Norwich and R. Nevin, *Il Nuovo Cimento*, 115B, 1137-1147, 2000.

†††. This section may have been expressed more tersely in the associated publication, “Sensory function in extraterrestrial beings”, K.H. Norwich and W. Wong, *Annales de la Fondation Louis de Broglie*, 22, No. 2, 161-168, 1997.

REFERENCES

- Aharonov, Y. and Bohm, D. 1961. Time in the quantum theory and the uncertainty relation for time and energy. *Physical Review*, **122**, 1649-1658.
- Bohm, D. 1989. *Quantum Theory*. Dover, New York. Published originally in 1951 by Prentice-Hall.
- Bohnenberger, J. 1981. Matched transfer characteristics of single units in a compound slit sense organ. *Journal of Comparative Physiology A*, **142**, 391-402.
- Broca, A. and Sulzer, D. 1902. *Journal de Physiologie et de Pathologie Générale*, **4**, 632-640.
- Chocholle, R. 1940. Variations des temps de réaction auditifs en fonction de l'intensité à diverses fréquences. *l'Année Psychologique*, **41**, 65-124.
- Evans, E.F. 1982. Basic physics and psychophysics of sound. In: *The Senses*, H.B. Barlow and J.D. Mollon Eds, Cambridge University Press.
- Gulick, W.L., Gescheider, G.A. and Frisina, R.D. 1989. *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*. Oxford Univ. Press, New York.
- Harris, J.D. 1963. Loudness discrimination. *The Journal of Speech and Hearing Disorders*, Monograph Supplement, **11**, 1-63.
- Lehmann, A.G.L. 1905. *Die körperlichen Äusserungen psychischer Zustände*. Part 3. *Elemente der Psychodynamik* (Translated by F. Bendixen). O.R. Reisland, Leipzig.
- Lindsay, P.H. and Norman, D.A. 1977. *Human Information Processing* (2nd edition), Academic Press, New York.
- Lochner, J.P.A. and Burger, J.F. 1961. Form of the loudness function in the presence of masking noise. *Journal of the Acoustical Society of America*, **33**, 1705-1707.
- Lochner, J.P.A. and Burger, J.F. 1962. Pure-tone loudness relations. *Journal of the Acoustical Society of America*, **34**, 576-581.
- Lockhead, G.R. 1992. Psychophysical scaling: Judgments of attributes or objects? *Behavioral and Brain Sciences*, **15**, 543-558.
- Marks, L.E. and Warner, E. 1991. The slippery context effect and critical bands. *Journal of Experimental Psychology: Human Perception and Performance*, **17**, 986-996.
- Meiselman, H.L., Bose, H.E. and Nykvist, W.F. 1972. Magnitude production and magnitude estimation of taste intensity. *Perception and Psychophysics*, **12**, 249-252.
- Norwich, K.H. and McConville (1991). An informational approach to sensory adaptation. *Journal of Comparative Physiology A*, **168**, 151-157.
- Nutting, P.G. (1907). The complete form of Fechner's law. *Bulletin of the Bureau of Standards*, **3**, 59-64.
- Riesz, R.R. 1928. Differential intensity sensitivity of the ear for pure tones. *Physical Review*, Series **2**, 867-875.
- Robinson, D.W. and Dadson, R.S. 1956. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, **7**, 166-181.
- Ueno, T. 1976. Luminance-duration relation in reaction time to spectral stimuli. *Vision Research*, **16**, 721-725.
- Wightman, F.L. and Green, D.M. 1974. The perception of pitch. *American Scientist*, **62**, 208-215.