Relativity in Human Judgment

Relativity in Human Judgment

By

Donald Laming

Cambridge Scholars Publishing

Relativity in Human Judgment

By Donald Laming

This book first published 2025

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-0364-4379-5
ISBN (Ebook): 978-1-0364-4380-1

The diagram on the front cover is adapted from Figure 3.1 on p. 22. This figure represents the principal experimental finding, from which all other theoretical results are derived.

CONTENTS

KEYWORDS

Absolute identification

Autocorrelation

'Bow' effect

Feedback

Limit to absolute identification

Marking examinations

Medical screening

Relative judgment

Sensory thresholds

Sequence effects

PREFACE

Suppose you are a schoolteacher signed up to mark GCSE (a public examination). You are given instructions and some initial training and then a bundle of scripts to mark. It is implicitly assumed that your marks do not depend on the order in which the different scripts are marked; but is that really so? Suppose there was some significant relation between the marks assigned to successive scripts; what would that relationship look like? These questions cannot be addressed with respect to your bundle of examination scripts, because the 'true' marks are unknown. But suppose the scripts are substituted with something else for which the true score is known exactly. In the psychological laboratory 1 kHz tones of different levels has been the most frequent choice, and the task is then called 'absolute identification'. Of course, results obtained with 1 kHz tones do not necessarily transpose to the marking of examination scripts. But, since schoolteachers engaged in marking scripts could equally serve as observers in the psychological laboratory, any difference must be to do with the different material being judged.

In 1951 it was discovered that observers are unable to identify more than five distinct levels of a 1 kHz tone absolutely–that is, identify without making comparison with any other tone. This discovery was not specific to loudness; it held for every sensory continuum that was examined. Observers endeavoured to identify increasing numbers of distinct stimuli, but errors multiplied as soon as that number exceeded 5. Why? In 1981 Christopher Poulton put a reprint under my nose that he had received from America[1]: "What do you make of this?" The experiment in this reprint asked observers to assign a number to each of a random series of 1 kHz tones according to their apparent loudnesses. This was my introduction to a fundamental relationship between successive judgments.

It seemed to me that there was no absolute judgment; all judgments are comparisons of one thing with another, so that the estimate of the loudness of a specific tone depends on the loudness of the tone that has preceded it.

---

[1] Baird et al, (1980)

Ordinarily, with everyday judgments, there is no way of establishing such an idea; but when someone has to make a long series of similar judgments –estimating the loudnesses of a long series of 1 kHz tones–it becomes possible to uncover the relationship between successive judgments. I then carried out a large experiment[2] (36,000 observations divided between four observers), again using 1 kHz tones, with a view to determining under what conditions that 'fundamental relationship' could be observed and to obtaining a sufficient volume of data to discover what it was. This book sets out a quantitative theory based on that data and, at the same time, reviews previous studies of absolute identification.

People make judgments everyday of all sorts of things and the next question is how far the theory set out here informs us about such judgments. For example, competitive sports divide roughly into three kinds: There are *games* like basketball, football and hockey, where opposing teams *score*; there are *races* where the winner is the '*first past the post*'; and there are *performances*, diving, figure skating, gymnastics, which are *assessed* by a panel of experienced judges. The theory provides scenarios that make some everyday judgments comprehensible, but such judgments rarely yield data that can be meaningfully analysed. Exceptions occur when someone has to make a long series of similar judgments, e.g. medical screening, marking public examinations, and inspection of baggage at an airport. Some of these 'outside the laboratory' judgments ae examined in the penultimate chapter.

[2] Laming (2023). The raw data from this experiment are available at https://www.researchgate.net/publication/367392647_Data7z

CHAPTER ONE

SOME RECENT HISTORY

In the 1950s it was discovered that observers were unable to identify more than 5 stimuli absolutely on any sensory continuum–that is, without comparison with some other stimulus. This discovery was not specific to any particular continuum; it was found to hold for every continuum tested. It is, therefore, not a property of the continuum, but of the process of judgment.

Garner (1962, p. 67) put this question:

> "If we were to use any perceptual or sensory continuum as a method of coding and displaying any information continuum (for example, suppose we want to indicate the altitude of an aircraft by the size of the symbol used to represent it?) is it better to use a small number of discrete steps on the [size] continuum, or a large number of small steps, or even to use the complete continuum with its infinity of steps?"

This practical question was addressed by Hake and Garner (1951), who asked observers to report the position of a pointer between two scale marks on an instrument dial. In different conditions of the experiment the pointer was set at random to one of 5, 10, 20 or 50 equally spaced sets of positions. Suppose the pointer indicated some physical variable of interest; how accurately is the value of that variable read by an observer? Hake and Garner found that maximum information throughput was achieved when the pointer was set to one of just 10 distinct positions; 20 and 50 distinct positions afforded no improvement. For practical purposes, then, observers were unable to distinguish more than 10 divisions of the interval between the two scale marks.

This was surprising and quite unexpected, because it has long been known that discrimination of visual separation is remarkably fine. Volkmann (1863) hung three threads, held taut by weights, from a bar in such a manner that the position of one thread could be finely adjusted by a screw. Those three threads provide an appropriate comparison with Hake and Garner's two scale marks and pointer; Volkmann's Weber fraction for discrimination of visual position was 0.008 (see Laming, 1986, Figs. 1.5 & 1.6, pp. 7 & 8).

The discovery that the number of stimuli that could be identified on a single continuum was severely limited seems to have spread rapidly; the experiment was soon replicated with many other continua. On most continua observers proved unable to identify more than 5 stimuli without reference to a second stimulus on the same continuum. The known exceptions–colour and angle of inclination–appear to contain natural anchors that serve as implicit references. Hake and Garner's observers identified 10 stimuli absolutely, but with respect to *two* scale marks–that is, 5 positions with respect to each mark. The existing experimental results as at that time were summarised by Miller (1956).

At that time much psychological theorising was dominated by 'information theory', specifically by intuitions arising from Shannon's theory of communication through a channel of limited capacity (Shannon & Weaver, 1949, see also Laming, 1968, p.5). Information transmitted is measured by the log likelihood-ratio statistic testing independence between the stimuli presented and the responses made (Kullback, 1959, p. 157). If logarithms are taken to base 2, the information is then measured in 'bits'. Limited capacity means simply that there is a limit to the *rate of transmission* of information through the communication channel. The analogous limit for absolute identification is *information per trial*.

A relevant comparison is the number of just noticeable differences that can be identified on the continuum in question. Visual separation has a Weber fraction of 0.008 (Volkmann, 1863, above), visual size (wavelength of a grating) 0.08 (Campbell et al., 1970), level of Gaussian noise 0.064 (Harris, 1950; see Laming, 1986, Table 5.1, pp. 76-77). How can absolute identification fail so dismally when discrimination is so fine?

Difference thresholds were traditionally measured using the Method of Constant Stimuli. On each trial a Standard stimulus was presented followed by a Comparison. Greater or less? Wever & Zener (1928) compared that procedure with their Method of Single Stimuli. They omitted the standard from each trial, with a 50% saving in stimulus presentations. Two experiments with lifted weights, demonstrated comparable precision with the Method of Constant Stimuli. This result was replicated by Pfaffmann (1935) using a salt solution (see Woodworth & Schlosberg, 1955, p. 219). The same precision could be achieved with 50% fewer stimulus presentations.

Sequential interactions

It has been known since Fernberger (1920) that successive psychophysical judgments are not strictly independent. Indeed, the accuracy achieved by the method of single stimuli, relative to the method of constant stimuli, suggests a relationship between each judgment and the preceding comparison stimuli that makes up for the missing standard. Holland & Lockhead (1968) reported an experiment designed to study this relationship. There were ten 0.5s 1200 Hz tones to be identified covering a range of 25dB. The stimuli were identified as ‘1’ … ‘10’, increasing with level; they were presented in random order. There were three observers, who were informed of the correct response at the end of each trial.

Figure 1.1 shows the average ‘error’ (that is, mean response less stimulus on a scale from 1 to 10; ex Holland & Lockhead, 1968, Fig. 3) on trial *n* contingent on the identity of the stimulus *k* trials previous, that is, on trial (*n*-*k*). Each data point represents 1200–1600 observations. The response on trial *n* assimilates (positive correlation) to the stimulus on trial *n*-1, but contrasts (negative correlation) to all preceding stimuli (trials *n*–2, etc.). That contrast appears to extend to 5 trials previous. This pattern of assimilation and contrast was replicated by Ward & Lockhead (1970), both with and without feedback, and (1971) in relation to previous responses.

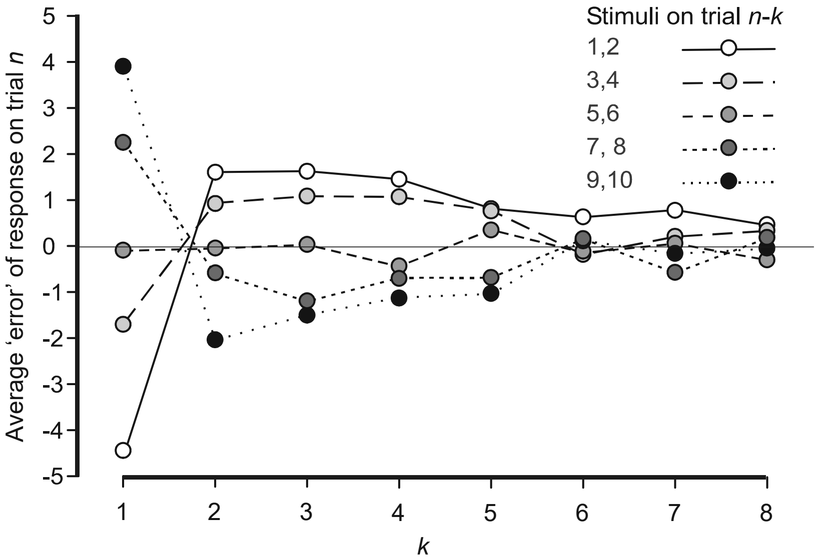

Figure 1.1. Average 'error' (mean response less stimulus) on trial *n* contingent on the stimulus *k* trials previous in the experiment by Holland and Lockhead (1968). Adapted from Holland and Lockhead (1968, Fig. 3).

And so to a study by Jesteadt et al. (1977) that contributes *two* fundamental discoveries:

(i) Four observers were asked to estimate the loudness of 27 1kHz tones of 500 ms duration ranging from 36 to 88 dB in 2 dB steps. The instructions copied Stevens' (1962) method of magnitude estimation, except that, instead of a cohort of observers who judged each stimulus twice only, each single observer contributed a total of approximately 1620 estimates, a volume of data sufficient to permit analysis of sequential relationships for each individual observer. Those relationships were analysed by linear regression of log response on the log trial stimulus and on preceding stimuli and responses up to 5 trials previous. Calculating multiple correlation coefficients for the immediate trial and then adding successive contributions from preceding trials (that is, calculating the multiple correlation coefficient successively for trials *n* to *n–k, k* = 0,…5), it appeared that while trial *n*–1 contributed materially to the multiple correlation, contributions from previous trials (i.e., *n*–2, *n*–3, etc.) were negligible (Jesteadt et al., 1977, Table 1). Sequential interactions appeared to be no more than trial-to-trial; that is, trial *n*-1 shapes the response on trial *n*; trial *n* then shapes the response on trial *n*+1, but independently of events on trial *n*–1 and preceding trials. "…the depth of the effects in the regular magnitude estimation experiments was much less than we had been led to believe." (Jesteadt et al., p. 95).

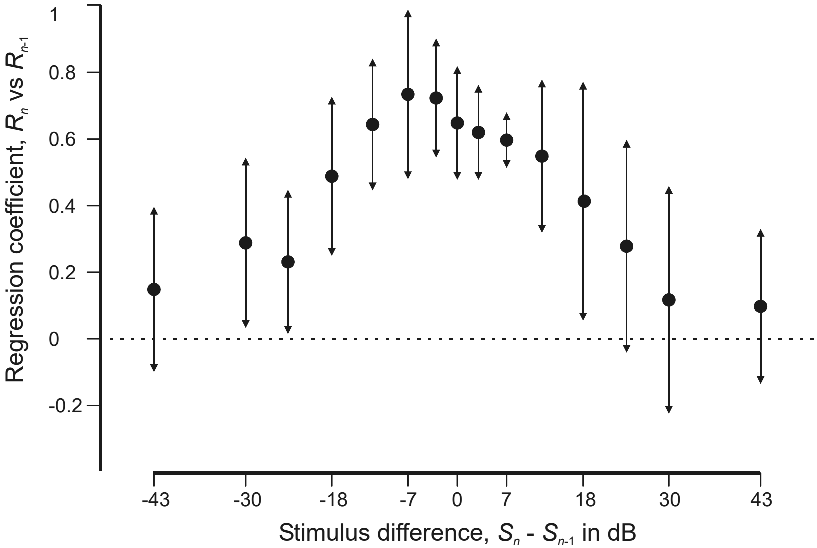

Figure 1.2. Mean partial regression coefficients, R_n vs R_{n-1}, in the experiment by Jesteadt et al. (1977). The different dB separations between successive stimuli have been collapsed into 15 groups of approximately equal numbers of trials. Each mean is the average of 8 observers; the arrows extend to ±1 standard deviation. Adapted from Jesteadt et al. (1977, Fig. 4).

Jesteadt et al.'s analysis *appears* at odds with Figure 1.1, but that appearance is illusory. Holland and Lockhead (1968) report the mean 'error' on trial *n* contingent on the stimulus *k* trials previous, without regard for the intervening trials (i.e., a *total* regression). Since the stimuli are presented in random order, that would appear to be unbiased. But if, as sketched above, the sequential interaction is simply trial-to-trial, then an effect of the stimulus on trial *n*-*k* on the 'error' on trial *n* can be transmitted through the intervening trials. Jesteadt et al., on the other hand, used multiple regression in which the 'error' on trial n is apportioned to those intervening trials. The contribution of trials prior to n-1 turns out to be negligible and this magnitude estimation experiment looks to be a Markov process.

(ii) The principal vehicle of sequential interaction is correlation between successive log responses; moreover, that correlation depends much on the dB difference between successive stimuli. Jesteadt et al. (1977, p. 100) calculated separate regression coefficients for different (dB) separations between successive stimuli. Figure 1.2 shows those regression coefficients, R_n vs R_{n-1}, grouped according to their dB differences. If a stimulus is repeated on successive trials, the successive responses are often the same, but when successive stimuli are different, there is much greater variation between the responses. This pattern of correlation has been replicated (Baird *et al.*, 1980; Green *et al.*, 1977; Luce & Green, 1978; Ward, 1979), though always with magnitude estimation/production procedures.

In the light of the experiment mentioned in the Preface (Baird et al, 1980):

> "I assert that, with rather few exceptions, human judgments of continuous sensory attributes are relative to the immediate context in which they are made." (Laming, 1984, p. 152).

A similar idea had previously been proposed by Holland and Lockhead (1968, p. 412):

> "In an absolute judgment task with feedback provided, it is proposed that Ss use the remembered magnitude of the stimulus on the preceding trial, and the numeric value of the feedback for that stimulus, as a standard for a comparative judgment of the presented stimulus."

But, in all the replications listed above, there was no feedback. An experiment was needed to compare magnitude estimation, on the one hand, with absolute identification without feedback on the other (Laming, 2023). If absolute identification failed to show the pattern of correlation exhibited in Figure 1.2, then that pattern must be specific to magnitude estimation; but if absolute identification *did* show that pattern, then that pattern must be characteristic of human judgement in general. It turns out (Chapter 3) that the pattern of autocorrelation revealed in Figure 1.2 is characteristic also of absolute judgment without feedback, and that the effect of feedback is, ironically, to attenuate that pattern.

This book presents an account of the principal phenomena of absolute identification/magnitude estimation from that point of view. More than a few experimental psychologists have attempted to model absolute identification starting from a variety of prior ideas about the internal representations of the stimuli and the underlying process of judgment; those different ideas are surveyed in Chapter 3. They invariably constrain the model that results. I choose instead to work from the pattern of autocorrelation in Figure 1.2, which is wholly empirical. The work by Holland and Lockhead (1968; Fig. 1.1) and by Jesteadt et al. (1977; Fig. 1.2) and many similar studies show that experimental data are everywhere confounded with sequential interactions, which obscure, or at the least diminish, underlying trends. An essential preliminary is to disentangle those sequential interactions.

CHAPTER TWO

THE LIMIT TO ABSOLUTE IDENTIFICATION

Suppose some number (N) of distinct stimuli on a single sensory continuum are to be identified. For 2, 3, and 4 stimuli (see Fig. 2.1 below), every identification is correct, but as N exceeds 5, errors of identification begin to appear. The number of errors does not increase *gradually* as N increases; instead, the information per trial reaches a limit equivalent to the identification of no more than five stimuli without error.

In more detail, on each trial of the experiment one of those N stimuli, selected at random, is presented and the observer asked to identify it. The data consist of the number of trials on which Stimulus i was presented and identified as Stimulus j. Now ask: Is the observer able to distinguish one stimulus from another or is he/she merely guessing 'in the dark'? Let n_{ij} be the number of trials on which Stimulus i was presented and identified as Stimulus j. Then, for N alternative stimuli,

$$ln\, \lambda = \sum_{i,j}^{N,N} n_{ij}\, ln\left(\frac{n_{ij} \sum_{i,j} n_{ij}}{\sum_i n_{ij} \sum_j n_{ij}}\right) \qquad (2.1)$$

is the log likelihood ratio statistic testing the independence of responses from stimuli (i.e. 'guessing in the dark'), and $-2\ln \lambda$ is distributed as χ^2 with $(N\text{-}1)^2$ d.f. (Kullback, 1959, p. 157). Taking logs to base 2 expresses the information throughput ($\ln \lambda$) in 'bits' (binary digits) and dividing by the number of trials gives a measure of how well the observer can distinguish the different stimuli. If all responses are correct, $\ln \lambda$ attains its maximum possible value ($\ln N$ per trial). Raised to the power of 2, this measure is equivalent to the number of different stimuli that can be distinguished without error.

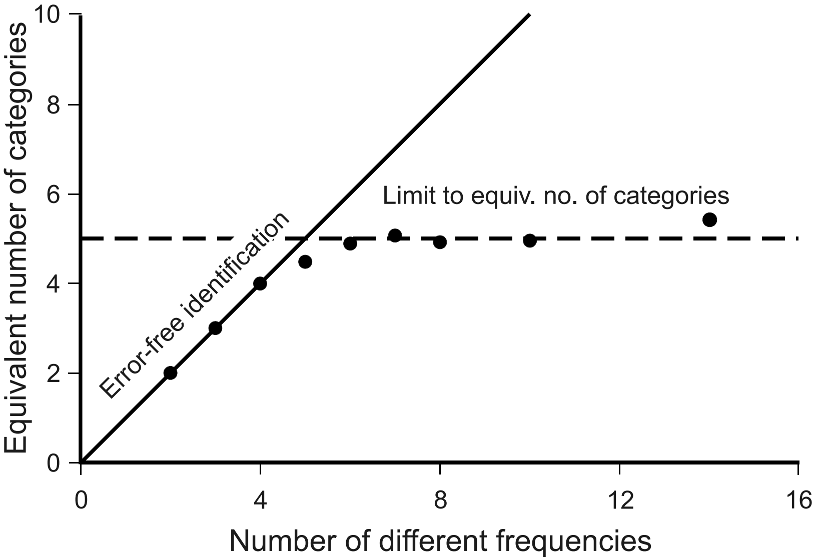

Figure 2.1. Accuracy of identification of auditory frequency in the range 100–8000 Hz. Figure from *Human Judgment: The Eye of the Beholder,* by D. Laming, p. 10, London: Thomson Learning, 2004. © Thomson Learning, 2004. Reproduced by permission. Data from Pollack (1952), averaged over six observers.

To illustrate, Figure 2.1 shows the equivalent number of stimuli identified in an experiment by Pollack (1952). An 85 dB pure tone was presented for 2.5 s, selected at random from some 2–14 different frequencies within the range 100–8000 Hz. Different numbers (*N*) of stimuli were presented in different blocks of trials. Observers were asked to identify the stimulus with a number between 1 for the lowest frequency and variously 2–14 for the highest. Following each response, the observer was told the correct answer. Some 25 s later another tone was presented, and so on. For 2, 3 and 4 alternative frequencies every identification was correct, but errors appeared as soon as the number of different frequencies reached 5 and increased through 6, 7, 8, 10, and, ultimately, 14. The increase in errors is not gradual; instead, as the number of alternative stimuli exceeds 5, errors increase rapidly until the information throughput reaches a limit equivalent to five.

This was surprising. Defining a 'just noticeable difference' as that difference in frequency that permits 75% correct identifications in a two-alternative forced-choice discrimination, it can be estimated from the data of Wier et al. (1977) that at 80 dB SL there are about 2000 such differences in the range 100–8000 Hz. But Pollack's stimuli were not presented in pairs as in a forced choice discrimination; instead, they were presented one at a time 25 s apart. In supplementary experiments Pollack, comparing the ranges 100–500 Hz and 100–8000 Hz, recorded limits equivalent to 3.5 and 4 stimuli respectively; and different distributions of eight frequency values within the range 100–8000 Hz gave resolutions equivalent to 3 to 4.3 stimuli. If the tones vary in loudness as well as frequency, observers can identify more than five distinct stimuli, though not so many as 25 (that is, not so many as five categories of frequency combined with five categories of sound level; Pollack, 1953). Indeed, the only experimental manipulation that improved identification was the provision of a reference tone preceding the tone to be identified. "This is tone no. 3". When tone no 3 was followed by itself, the accuracy of identification was equivalent to 10.6 stimuli, but that improvement did not extrapolate to adjacent stimuli. There is a dramatic difference of two orders of magnitude between the identification of a single frequency and the discrimination between two fixed frequencies in close temporal proximity.

Following the initial study by Hake & Garner (1951), other sensory continua were quickly examined. The results are set out in Table 2.1. 'Bits/trial' in Column 4 is the average information throughput. Column 5 expresses that throughput as an equivalent number of stimuli identified without error.

The equivalent number of stimuli that can be identified without error does not exceed 5, except for colour, where it is arguable that there is a natural reference point in the yellow; and angle of inclination (Muller, Sidorsky, Slivinsky, Alluisi & Fitts, 1955). This last study was conducted in a darkened cinema auditorium where, arguably, the vertical and horizontal divided the range of inclinations into four. Of course, Hake and Garner's (1951) observers distinguished 10 positions of the stimulus pointer, but they had *two* scale marks to which reference could be made and were able to fractionate between those two marks. Unless there is some physical reference like a ruler or a scale pan against which comparison can be made, five categories of judgment is the pragmatic limit.

Table 2.1. Limits to absolute identification of different stimulus attributes.

Sensory modality	Attribute	Source	Bits/trial	Equivalent no of stimuli
Audition	frequency	Hartman (1954)	2.3	4.9
		Pollack (1952)	2.3	4.9
		Siegel (1972)	1.6	3.0
	intensity	Braida & Durlach (1972; from calculations by Marley & Cook, 1984)	1.9	3.7
		Garner (1953)	2.2	4.6
		Norwich, Wong & Sagi (1998)	2.2	4.6
Colour	Munsell hues	Eriksen & Hake (1955b)	3.1	8.6
		Conover (1959)	3.5	11.3
	Spectral hues	Chapanis & Halsey (1956)	3.3	9.8
Smell	Odor intensity	Engen & Pfaffmann (1959)	1.5	2.8
Taste	Salt	Beebe-Center, Rogers & O'Connell (1955)	1.7	3.2
	Sweetness	Beebe-Center, Rogers & O'Connell (1955)	1.7	3.2
Touch	Cutaneous electric current	Hawkes & Warm (1960)	1.7	3.2
Vision	Angle of inclination	Muller, Sidorsky, Slivinsky, Alluisi & Fitts (1955)	4.5	22.6
		Pollack (ex Miller, 1956)	2.8/3.3	7.0/9.8
	Area in general	Pollack (ex Miller, 1956)	2.6/2.7	6.1/6.5
	Area of a circle	Alluisi & Sidorsky (1958)	2.7	6.5
	Area of a square	Eriksen & Hake (1955a)	2.0	4.0
		Eriksen & Hake (1955b)	2.8	7.0

Table 2.1. Limits to absolute identification of different stimulus attributes (cont)

Sensory modality	Attribute	Source	Bits/trial	Equivalent no of stimuli
	Area of complex figure	Baird, Romer & Stein (1970)	2.1	4.3
	length	Baird, Romer & Stein (1970)	2.4	5.3
		Pollack (ex Miller, 1956)	2.6/3.0	6.1/8.0
	location	Coonan & Klemmer (ex Miller, 1956)	3.2/3.9	.2/14.9
		Hake & Garner (1951)	3.2	9.2
	Luminance	Eriksen & Hake (1955b)	2.3	4.9

Entries with a slash indicate different values for short and longer exposures of the stimuli.
Results collated from Miller (1956), Garner (1962), Laming (1984) and Stewart et al. (2005).

Why the fuss?

The comparison between the limit to accuracy of identification of single frequencies in Figure 2.1 (at most 5) and the 2000 'just noticeable' differences in the range 100–8000 Hz (Wier et al., 1977) exemplifies the problem. Table 2.2 lists Weber fractions for other continua examined in Table 2.1. For some continua (e.g. visual separation, Weber fraction 0.008, ex Volkmann, 1863; see Laming, 1986, Fig. 1.5, p. 7) discrimination is especially acute. So, how can absolute identification fail so dismally when threshold discrimination is so fine?

Thurstone's 'Law of Comparative Judgment'

Thurstone (1927b) proposed that a stimulus, as perceived by the observer, is intrinsically variable. This would explain why psychophysical judgments, notably threshold discriminations between different stimulus magnitudes, are variable. Taking that variability to be normally distributed, Thurstone sought to establish a stimulus scale from experimental data. As an example,

Table 2.2. Weber fractions (75% correct, 2-alternative forced choice) and psychophysical method for the different stimulus attributes in Table 2.1.

Sensory modality	Attribute	Source of data
Audition	Amplitude of Gaussian noise	Harris (1950)
	Amplitude of a pure tone	Jesteadt et al. (1977)
Colour	Hue	Indow & Stevens (1966)
	(a) Red-yellow	
	(b) Green-yellow	
	Saturation	Panek & Stevens (1966)
Contrast	Sinusoidal grating	Kulikowski (1976)
	1 cd/m²	
	10 cd/m²	
Smell		Stone (1963) Stone & Bosley (1965) Stone et al (1962)
Taste		Schutz & Pilgrim (1957)
	(a) Caffeine	
	(b) Citric acid	
	(c) Common salt	
	(d) Sucrose	
Touch	Amplitude of vibrotactile stimulation	Craig (1974)
	Frequency of vibrotactile stimulation	Mountcastle et al. (1969)
	Pressure on the skin	Gatti & Dodge (1929)
Vision	Luminance	Cornsweet & Pinsker (1965)
	Visual separation	Volkmann (1863)
	Visual size (Wavelength of grating)	Campbell et al (1970)

† At 80 dB SL

§ $0 < C < 1$

* Yes/No signal detection with 2 stimulus presentations, $< f, f >$ or $< f, f+\Delta f >$.

‡ Lower difference threshold.

¶ For several different substances, including acetic acid, with similar odours.

Table 2.2. Weber fractions (75% correct, 2-alternative forced choice) and psychophysical method for the different stimulus attributes in Table 2.1.

Weber fraction	Threshold criterion & psychophysical method	Equivalent Weber fraction
0.064	0.75, Constant Stimuli	0.064
$0.23A^{-0.14}$†	0.71, 2AFC staircase	0.077†
	S.D., Average Error	
0.006		0.004
0.029		0.020
0.025	0.75, Single Stimuli	0.018
	0.75, Constant Stimuli	
$0.054C^{-0.29}$§		$0.054C^{-0.29}$§
$0.038C^{-0.29}$§		$0.038C^{-0.29}$§
0.250	0.84, Constant Stimuli	0.170
	0.75, Single Stimuli	
0.302		0.214
0.224		0.158
0.153		0.108
0.172		0.122
0.245	0.75, 2AFC staircase	0.245
0.100	S.D., Yes/No sig.det.*	0.135
0.171	Limits	0.121
0.170	0.71, 2AFC, staircase	0.210
0.008	P.E., Average Error	0.008
0.08	S.D., Yes/No sig.det.	0.038

† At 80 dB SL

§ $0 < C < 1$

* Yes/No signal detection with 2 stimulus presentations, $< f, f >$ or $< f, f+\Delta f >$.

‡ Lower difference threshold.

¶ For several different substances, including acetic acid, with similar odours.

Garner & Hake (1951) constructed such a scale from judgments of loudness. Thurstone (1927a) then distinguished five different formulations of this idea, of which ‘Case V’ is much better known these days as the normal, equal variance, signal-detection model.

Torgerson (1958) developed Thurstone’s ‘Case V’ to accommodate a linear array of stimuli. There are N normal distributions of means μ_i and common variance σ^2,

$$f_i(x) = (2\pi\sigma^2)^{-1/2}\exp\{-(x-\mu_i)^2/2\sigma^2\},\ i = 1,\ldots, N, \quad (2.2)$$

one for each stimulus value, $i = 1,\ldots, N$, and N-1 criteria, c_j, $j = 1,\ldots, N$-1. The probability of response j given stimulus i, $P(R_j|S_i)$, is represented by the proportion of the S_i distribution that lies between criteria c_{j-1} and c_j;

$$P(R_j|S_i) = \int_{c_{j-1}}^{c_j} f_i(x)\ dx. \quad (2.3)$$

In practical applications the standard deviation is commonly fixed at 1 and the mean of the smallest stimulus at 0; all the other stimulus means and criterion values are then estimated from the data. This implicitly establishes an interval scale of the stimulus continuum in units of the standard deviation and provides a scaffold for the absolute identification of whatever stimulus is presented, subject, of course, to the error σ^2. Equation 2.2 has provided the foundation for many studies of absolute identification, notably Durlach & Braida (1969), Braida & Durlach (1972), Treisman (1985) and Luce et al. (1976).

Suppose the variance in Equation 2.2 be chosen to define a ‘just noticeable difference’. Then the sequence of means μ_i would mark out successive JNDs. Increasing the range covered by a fixed number of stimuli, that is, spacing the stimuli more widely, would lead to continually improved identification. But it does not happen so. Braida and Durlach (1972, Expt. 4), in a much-quoted experiment, compared absolute identification with feedback for ten 500ms bursts of 1000 Hz tones spaced, in different blocks of trials, at 0.25, 0.5, 1 2, 3, 4, 5 and 6 dB intervals (and covering successive ranges of 2.25, 4.5, 9, 18, 27, 36, 45 and 54 dB) with the highest level always at 86 dB. There were about 1875 trials per condition for each of three observers. Figure 2.2 shows the information throughput, expressed as bits/trial, for each of three observers. The information throughput increases as expected for the 3 closest spacings, but then reaches a limit. Figure 2.2 and the results listed in Table 2.1 are at variance with Equation 2.2. Something has to give. The experimental results can be replicated. It has to be Equation 2.2.

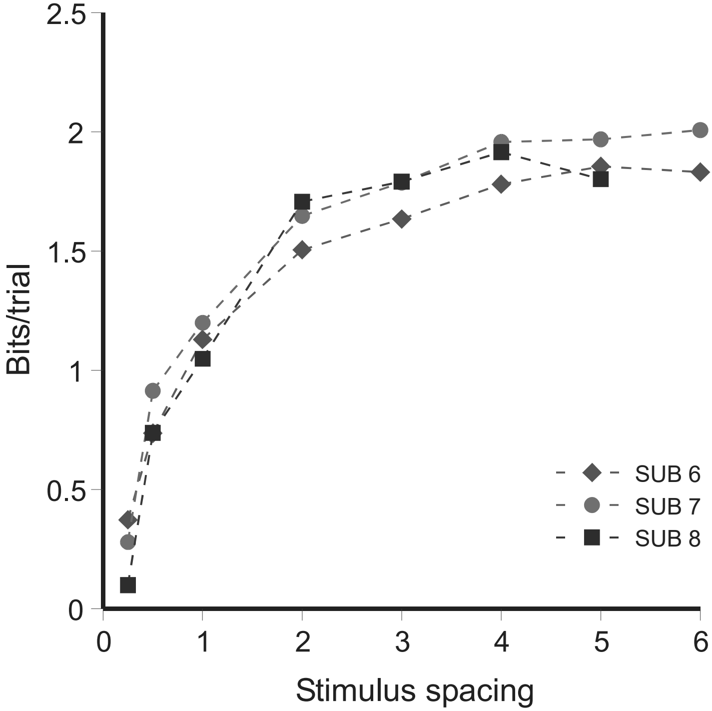

Figure 2.2. Information transmitted per trial for three observers in Braida and Durlach's (1972) Experiment 4. Figure adapted from Laming (2023, Fig. 11), *Autocorrelation in category judgement*. Quarterly Journal of Experimental Psychology, 76(12), 2869, © Experimental Psychology Society 2023.