CONTEXT AND JUDGMENT

Gregory R. Lockhead
Duke University, Durham, NC, USA

Abstract

Fechner's psychophysical techniques and their extensions allow measuring several psychological processes. Examples in this tutorial demonstrate the value of these measures for studying sequential processes associated with univariate and bivariate psychophysical judgments, for examining simultaneous context effects and stimulus set effects in classification, and for measuring the amounts by which variations of nominally irrelevant features affect judgments. Psychophysical tasks are rich in psychological processes and these various measures help to show ways that choice depends on memory, on comparative behavior, and on response strategy.

E. H. Weber is credited with uncovering two stunning regularities in human performance: The amount by which a intensity must change in order to be detected is proportional to that intensity, and different domains have different proportionality constants. Fechner called this "Weber's Law" and built upon it to create psychophysics, produce Fechner's Law, and develop a foundation for scientific psychology. Fechner's Law, and arguments offered to replace it, all demonstrate that judgments of attribute intensity, when averaged across trials, increase lawfully with stimulus intensity. This finding produced valuable scales for industrial use and provided a foundation for 150 years of psychophysical research.

Yet, we still do not know how these functions are produced. A common interpretation is people abstract the attribute of interest and judge its magnitude independent of other features of the stimulus and of the situation (Stevens, 1975). Accordingly, except for unexplained noise, psychophysical judgments are independent of preceding events, alternative potential stimuli, and irrelevant features. By this account, when people are asked to report, for example, the brightness of a stimulus that has a color, size, shape, location, and history, as well as an intensity, they do the task by abstracting the intensity [the brightness] and judging it directly.

However, people usually cannot judge attributes independent of context and so this classic interpretation does not describe what they do, or at least does not describe everything they do. Which of these two theoretical possibilities is the case, if either is, is not addressed here (cf. Lockhead, 1992). This paper reviews some evidence for three classes of context effects, notes that psychophysical methods are useful to measure them, and suggests that such measures can indicate the sources of the effects. The contextual factors considered in this brief report are relations among the stimuli being judged, variations of irrelevant attributes, and event sequences. Any full review of context effects would include many more factors.

Sequence effects in absolute judgments and magnitude estimations.

In magnitude estimation tasks, people are presented several stimuli, one at a time, and are asked to judge the magnitude of some feature of each stimulus. The stimuli might be tones, and loudness might be judged. The first tone is assigned some number, such as 100. If the next tone appears twice as loud, it is to be called 200, if it appears a fourth as loud, it should be labeled 50. Perhaps it is called 200. The response to the next tone is then produced by comparing it with this 200 magnitude tone. And so on for many trials of randomly presented tones. The responses to each stimulus tone then are averaged across trials, and the common finding is that response magnitude increases linearly with stimulus intensity expressed in decibels (e.g., Stevens, 1975).

In such experiments, the observer is asked to do these things: 1) abstract the relevant attribute, 2) judge its magnitude, 3) recall the prior magnitude, 4) divide these two magnitudes, 5) recall the prior response, 6) multiply the outcomes of steps 4 and 5, and 7) respond. This is a complex process with many opportunities for error, and there are regularities in the data that are not expected on the basis of the instructions. One such regularity is each stimulus is reported as being overly similar to the previous stimulus, as if there is some magnetic attraction between successive trials. This effect, known as assimilation, also occurs in absolute judgment tasks. A second effect, known as contrast (perhaps like magnetic repulsion), also occurs between the response and earlier trials. Contrast is indicated when the stimulus is judged as overly different from stimuli several trials back in the sequence and is particularly marked when feedback is given after each response.

We know that assimilation and contrast are not simply sensory effects because stimuli are not needed for them to occur. They occur when the stimulus generator is turned off and observers guess what stimulus should have been presented. They also occur when the subject guesses the stimulus extra sensory perception studies. In these last two cases, accuracy is at chance but the magnitudes of assimilation and contrast are large. Thus, response processes are involved in producing these context effects. A variety of sequence effects is summarized in Lockhead (1984).

While responses are important, stimuli are also involved in sequence effects. This is shown a successive ratios judgment task. This is a magnitude estimation procedure except steps 5 and 6 above are omitted. The observer compares each stimulus with the memory of the previous stimulus and reports the ratio between them. Of several context effects found in such data (Lockhead & King, 1983), a prominent one is seen on trials when the stimulus repeats. Then, the ratio between successive stimuli is one and so the response should be one. But "1" one is almost never given in these cases, even though "1" is frequently given when the stimuli are physically different. Consider a successive ratios study of tones when loudness is judged. On trials that the tone two trials ago, trial S_{N-2}, was quieter than the subsequent two identical tones, no matter what their intensity, the typical response is greater than one. That is, S_N is then judged as louder than the identical $S_{N-1.}$ But on trials when S_{N-2} was louder than two subsequent identical tones, the typical response is less than one. Then, S_N is judged as quieter than the identical S_{N-1}. This result is expected if each intensity assimilates in memory toward the tone that preceded it, and if the current stimulus is compared with this distorted memory. This result is not expected in the absence of context effects.

Attribute relations in classification.

In typical psychophysical tasks, stimuli only differ along the dimension being judged. This means that contextual effects associated with other dimensions of the stimuli, if there are any, cannot be seen in the data. However, when the stimuli differ randomly from trial to trail on two dimensions, and again are classified on the basis of only one dimension, this variation in the "irrelevant" dimension of the stimulus often affects performance. This finding allows the idea that other dimensions than the one judged may be involved in all psychophysical tasks, that the entire stimulus object is involved, even though this cannot be known in the typical experiment where there is no variation of anything except the dimension of interest to the experimenter. The simplest case to examine this uses two levels of a relevant dimension and two levels of an irrelevant dimension. For example, when the stimuli in such a 2X2 matrix are tones that vary randomly in loudness (quiet and loud) and in pitch (high and low), and when loudness is judged while pitch is irrelevant, performance is poorer (longer response times and more errors) than in the equivalent task except that pitch does not vary (Garner, 1974, summarizes many such studies).

When this stimulus matrix is made larger than 2X2, such that more than two stimuli and two responses are involved, the observed context effects help indicate what is involved in the subjects' behaviors. One such study used six levels of rectangle height, six levels of rectangle width, and six responses. In one such condition, height (H) and width (W) were linearly correlated (H1 was paired with W1, H2 with W2, 3 with 3, 4 with 4, 5 with 5, and 6 with 6). In another condition, called sawtooth correlated, H and W were again perfectly correlated but these pairings were 1-1, 2-3, 3-5, 4-2, 5-4, and 6-6. The task in both conditions was to classify the stimuli according to width. If people do simply judge width, then performance should be the same in both conditions . This is because the widths are the same. However, responses were faster and more accurate for the sawtooth set than for the linear set. Particularly relevant concerning the importance of context is that stimuli 1-1 and 6-6, which are the identical objects in the two sets, are classified faster and more accurately in the sawtooth set than in the linearly paired set (Monahan & Lockhead, 1977).

This set effect also occurred in a study with 20 values of a single dimension (line tilt) and with 20 responses. When a single one line could take 20 orientations to be judged in an absolute identification study, classifications were 32% correct. When four such lines were arranged in the configuration of a face (2 eyes, nose, and mouth) and were linearly correlated (each line had the same tilt at each position), performance improved to 37% correct. But when the same 20 line orientations arranged as a face were sawtooth correlated, performance was 100% correct (Lockhead, 1970).

Range of variation of an irrelevant attribute.

In univariate tasks where the stimuli differ on only the dimension being judged, performance depends on range. When the stimuli vary over a large range, such as from very dim to very bright lights, brightness judgments measured in stimulus intensity units are less precise than when the stimuli vary over a smaller range, such as from dim to not quite so dim (Parducci & Perrett, 1971; Gravetter & Lockhead, 1973). Too, stimulus sequence affects performance in such univariate tasks (Bertelson, 1961; Lockhead 1984).

In bivariate tasks, where stimuli differ randomly on two dimensions but only one dimension is judged while the other is irrelevant, classification additionally depends on the range of the irrelevant dimension. Performance is poorer when the range of the irrelevant dimension is larger (Lockhead, 1992, Fig. 6). Sequence also affects performance in bivariate tasks (Felfoldy, 1974).

Since the average magnitudes of trial-to-trial stimulus differences are necessarily larger in conditions where stimulus range is larger, these separate findings suggest that irrelevant sequence and irrelevant range both might affect performance within a single task. Huettel & Lockhead (1999) examined this in bivariate classification studies and one of the results is summarized here. In Experiment 3A of that report, observers classified tones according to loudness. The stimuli had two levels of this relevant dimension (76 and 80 dB) and 12 levels of the irrelevant dimension of frequency or pitch (12 MIDI notes from middle C to B: 523 Hz to 987 Hz). One of these 24 stimuli was randomly selected for each of many trials, and the observer pressed one of two keys to indicate its loudness. Pitch was uninformative.

This stimulus set provides two trial-to-trial sequences on the relevant dimension of loudness; repeat the intensity (and thus repeat the response) and change the intensity (and thus change the response). This set also provides 12 sequences on the irrelevant dimension of pitch; repeat the pitch and change the pitch by 1 step (to an adjacent value) up to 11 steps (from the highest to the lowest frequency, or vice versa). Since pitch and loudness were randomized, on half the trials these pitch differences were associated with a loudness change, and on the remaining trials loudness did not change from the previous stimulus. Ignoring the few errors, the average effects of these sequences on response times are shown in Table 1 (from Huettel & Lockhead, 1999, Figure 8 and Table 3).

Table 1.

Average response times in msec to report the loudness of a stimulus when it repeated from the prior trial (top row) and when the loudness changed from the prior trial (bottom row), for each of the eleven frequency (pitch) differences.

Loudness repeats:	502	521	555	571	584	591	602	610	638	612	658	645
Loudness changes:	588	560	560	568	568	550	563	564	569	571	528	579
Semitone change:	0	1	2	3	4	5	6	7	8	9	10	11

On trials when pitch repeats (leftmost data column), responses are faster when loudness also repeats than when loudness changes (502 vs. 588 ms). Since the response repeats when loudness repeats and changes when loudness changes, this result is consistent with Bertelson's (1961) report of a response repetition effect in univariate tasks.

On trials when loudness repeats and pitch changes (top data row), the response also repeats, and response times increase from 502 to 645 ms. with increasing pitch differences. However, when the response changes, and so loudness also changes (bottom data row), responses are independent of pitch changes. This reliable interaction between effects of relevant and irrelevant stimulus change between trials indicates a complex process in this nominally simple psychophysical task.

Object Constancy, Context, and Reading Disability:

Many children in the Western world are characterized as having dyslexia. When most of these children are shown the letter form "q", they do not know if it should be called "pea", "dee", "bee", or "queue." This difficulty is not because of phonetic confusions, "queue" sounds nothing like "pea." It is also not because these children cannot read letters. I have never encountered a seeing child who cannot identify "X" and "O." The difficulty occurs because p, d, b, and q are similar to one another in a particular way. When rotated or reflected, these forms are essentially identical. Consistently, such children also tend to confuse M with W, n with u, and S with Z and 2.

One source of this reading difficulty may involve object constancy. This is the perceptual skill that allows objects to remain perceptually invariant across rotations. For example, a pet dog is seen as the same dog whether facing left or turned around. Indeed, imagine the turmoil if objects appeared to be different things when viewed from different orientations. Object constancy is apparently available from birth and, just as for other objects, allows a letter to be seen as the same thing independent of its orientation. Thus, at least until the needed learning or maturation occurs in the child, q really is d, and it is also p and b. It is no more appropriate to expect untrained children to discriminate these letters than to expect them to discriminate their dog who has turned around as being a different thing. By to this argument, all children must learn to break object constancy for letters if they are to learn to read.

A different argument is these children are letter-blind analogous to people who are color-blind. If this is so, then maturation will not occur and training will not be effective and the children will not learn these discriminations. In this case, just as we would not ask a color blind person to discriminate red from green, we ought not ask a letter blind person to discriminate b from d. Concerning this thesis, it is notable that colorblind people are allowed to drive automobiles even though they must discriminate red from green traffic lights. How is this done? The answer is not that they have learned the colors. It also is not that "red" is on top in the traffic light box; location information is not available at night, and in some regions of the world the lights are arrayed horizontally. Rather, the answer is that the traffic lights are not quite red and green. The "green" is sufficiently shifted toward "blue" that color blind people distinguish it from "red." Since many dominant wavelengths are classified as red and many others as green, color-normal people are not aware of this small modification. And color-blinds are not particularly confused by the names other people use to describe the lights since color names regularly confuse them. They only need to know when to stop and when to go, and a slight wavelength modification "cures" them in this regard.

Perhaps the alphabet can be similarly "fixed" to “cure” letter blind people, no matter why they have these confusions. To examine this, the normal letters b, d, p, and q were modified to make them more distinctive (Lockhead & Crist, 1980). A dot was added inside the loop of the d, an accent was added to the middle of the bar of the q, and another accent was added to the top of the bar of the b; p was not changed. When these modified letters are rotated or reflected, none of these “distinctive” letters produces another letter in the font. In one set of studies using such letters, children sorted normal letters that were printed in sans serif font, and also sorted these distinctive letters. In different conditions, the sorting bins were labeled with normal letters or with distinctive letters, and the cards being sorted were all normal letters or all distinctive letters. The subjects in three studies were college students, reading-normal children, and reading-disabled children.

The distinctive letters were consistently sorted faster and with fewer errors than the normal letters. Indeed, first grade dyslexic children sorted the distinctive letters, which were novel, faster and more accurately than second grade, reading normal children sorted the normal letters, with which they has considerable experience. This is at least a *one year advantage for “dyslexic” children over normal children.* At least for these classification tasks, reading disabled children are not disabled when using this distinctive font. Furthermore, the letter p, which was not modified, is easy for dyslexic children to "identify" or classify when it is used in the distinctive set, but the identical form is difficult for them to classify when it us used in the normal set. Thus, p (and by extension b, d, q, M, W, n, u, etc.) is not difficult for “dyslexic” children to process. Rather, p is difficult for them to discriminate when it appears like other letters in the alphabet and so is confused. “p” is easy to classify in the distinctive set where it is perceptually unique. This context effect suggests that less is wrong with the children and more is wrong with the alphabet than we have supposed.

Conclusion:

People do not directly identify stimuli or their attributes. They compare each stimulus event in memory with known possibilities, and judge it in comparison to perceived relations and remembered alternatives, all of which involve sequential and simultaneous comparison processes. Psychophysical judgment is a rich and complex task that we are coming to better understand through various uses of psychophysical measurement.

References:

Bertelson, P. (1961) Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology, 13,* 90-102.

Garner, W. R. (1974) *The Processing of Information and Structure*, Maryland: Erlbaum.

Gravetter, F., & Lockhead, G. R. (1973) Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80,* 203-216.

Huettel, S. A, and Lockhead, G. R. (1999) Range effects of an irrelevant dimension on classification. *Perception & Psychophysics, 61*, 1624-1645.

Lockhead, G. R. (1970) Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology, 85*, 1-10.

Lockhead, G. R. (1984) Sequential predictors of choice in psychophysical tasks. In S. Kornblum and J. Requin (Eds.), *Preparatory States and Processes*, Hillsdale, N.J., Erlbaum.

Lockhead, G. R. (1992) Psychophysical scaling: Judgments of attributes or objects? *Behavioral and Brain Sciences, 15*, 543-558.

Lockhead, G. R. & Crist, W. (1980) Making letters distinctive. *Journal of Educational Psychology, 72,* 483-493.

Lockhead, G. R. & King, M. C. (1983) A memory model of sequential scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 461-73.

Monahan, J. S., Lockhead, G. R. (1977) Identification of Integral Stimuli. *Journal of Experimental Psychology:General, 106*, 94-110.

Parducci, A. & Perrett, L. F. (1971) Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology, 89*, 427-52.

Stevens, S. S. (1975) *Psychophysics: Introduction to its perceptual, neural, and social prospects.* ed. G. Stevens. N.Y, Wiley.