# Absolute Judgments Are Relative: A Reinterpretation of Some Psychophysical Ideas

Gregory R. Lockhead
Duke University

A central theoretical assumption in classical psychophysics is that people judge the intensities of stimulus elements; for example, observers directly report the loudness of a tone or the intensity of a shock. A methodological assumption in classical psychophysics is that averaged data demonstrate this theoretical view. It is shown in this article that both assumptions are wrong and that the psychophysical laws of Weber, Fechner, and Stevens are not general. Rather, psychophysical judgments are made in relation to contexts and memories, measures of which provide new information about psychophysical judgments and new understandings of channel capacity, the local–global distinction, and the source of noise in signal detection theory.

Ernst Weber (1846/1965) showed that the amount by which the intensity, $I$, of a stimulus feature, such as its weight, must change for a difference to be perceived is a fixed proportion, $k$, of its magnitude, that is, $\Delta I/I = k$. This is referred to as the *just noticeable difference* and is known as Weber's law. On the basis of this finding, Gustav Fechner proposed a ruler to measure psychological magnitudes that he published in 1860 in his *Elements of Psychophysics*. His insight became an important basis for the creation of psychology as a science because it promised scales for measuring psychological events. His resulting law is that the psychological magnitude, $R$, of a stimulus element, $I$, equals the logarithm of its physical magnitude multiplied by a constant, $k$: $R = kLogI$. Nearly 100 years later, S. S. Stevens (1961) suggested replacing this law with his own law, $R = aI^{\beta}$. Other authors have suggested related equations, all of which make the same theoretical and procedural assumptions discussed subsequently (cf. Luce & Suppes's, 2002, review), and so they are considered together.

There are at least two excellent reasons for continued support of these laws: They conform to intuitively appealing theoretical ideas, and their expressions excellently describe thousands of data sets. Nonetheless, the laws are not general across environments, and the measures used to support them mask what people actually do when they attempt to judge an element of a stimulus event. Thus, perhaps extending the work of Stevens, who wrote, "To honor Fechner and repeal his law" (1961, p. 80), one purpose of the present article is to honor Weber, Fechner, and Stevens and to repeal their laws.

First, however, it should be noted that data supporting the classic theories are easy to obtain and are compelling. For example, to examine Stevens's law using the method of magnitude estimation, which Stevens created for this purpose, you can select several values along some intensive dimension. You might use six light intensities. On each of many trials, randomly select one of the lights and present it separately on a fixed background. For the first light, tell an observer to label the intensity "100." Then turn that light off and present the next randomly selected light, and ask the observer to judge its brightness relative to the "100" light with the instruction that "if it appears half as bright it should be labeled 50, and if it appears three times as bright it should be called 300." Then ask the observer to judge a third light in relation to the second light, a fourth in relation to the third, and so forth, for many trials. You might collect about 30 judgments for each of the six randomly presented lights.

Now, plot the average response to each light against the logarithm of its physical intensity and connect the dots. I can confidently predict the result is such that a best-fitting straight line

Correspondence concerning this article should be addressed to Gregory R. Lockhead, Department of Psychological and Brain Sciences, Duke University, Durham, NC 27708-0086. E-mail: greg@psych.duke.edu

will be consistent with Stevens's law, $R = aI^\beta$. This is a good classroom or laboratory exercise, because the results are so regular they seem almost eerie to beginning students. How, then, could one not accept Stevens's law?

Other measures are used to test the other two classic laws, wherein again the outcomes usually conform compellingly to the equations. Many such findings are reported in the literature as consistent with the underlying thesis in Fechner's *Elements of Psychophysics*, which is that people abstract and judge the intensity of some element of a stimulus event. Nonetheless, two difficulties with all such studies call this interpretation into question.

One difficulty with the classic theories is the assumption that people judge the physical intensity of an attribute of the stimulus. For the light intensity example, the assumption is that luminous intensity determines brightness. Accordingly, each response is defined as a match to the light's intensity adjusted by different (usually) constants in each case: $R = aI^\beta$ for Stevens's law, $R = kLogI$ for Fechner's law, and $\Delta I/I = k$ for Weber's law.

However, the intensity of the light itself seems never to be directly available for judgment. As simultaneous contrast shows, the brightness of any light depends on its environment; the same light appears black, gray, white, or colored depending on its surroundings. Accordingly, if the brightness study is conducted again but with a different background, the data will be different. Indeed, they will be markedly different if some of the judged lights are less intense than the background and others are more intense. Hence, it cannot be just the light that is judged. Furthermore, image stabilization research shows that each light quickly disappears if it does not change over time at the receptors. Thus, again, intensity is not a sufficient definition of the stimulus.

A more correct definition of the stimulus is that the perception of the brightness of a light is a function of energy differences over time and space. Therefore, a more correct definition of the processes involved in responding to that stimulus is that judgments are based on *relations* between the stimulus and its context rather than on the intensity of an element or attribute of the stimulus (Lockhead, 1992).

A second difficulty with the classic presentations is the incorrect assumption that responses do not depend importantly on factors that are seemingly controlled. Three such confounding factors are considered briefly here: stimulus range, response processes, and event sequences.

Concerning sequences, and again using brightness as the example, consider the case in which a light repeats on successive trials during a psychophysical scaling study. If there are five different stimuli in the task, such repetitions will occur on about 20% of the trials. If judgments are independent of sequence, the repeating lights should be perceived and reported as equal because they are in fact identical. But the two stimuli are not usually judged the same. Rather, the second of these two lights tends to be judged brighter than the first one if the light that just preceded these two was dim, but this second light tends to be judged dimmer than the first one when the light two trials earlier was relatively intense or bright. The first interpretation to account for this result was that the first light assimilated toward what preceded it (and, not discussed here, as contrasted to earlier events; Holland & Lockhead, 1968). This interpretation was later replaced with the suggestion that the perception or memory of the first light assimilates toward the already-biased memory of what preceded it, and the response to the second light is partially determined by the earlier events (Lockhead & King, 1983, Equations 8 and 9; see also interpretations by Braida & Durlach, 1988; Laming, 1997; Luce, Green, & Weber, 1976; and Treisman & Williams, 1984).

Although sequence effects have been widely studied, the issue has been largely ignored in terms of its effect on psychophysical scaling. Apparently, this is for two reasons. First, these effects are extraneous to the theory pursued. Perhaps, just as air interferes with measures of the gravitational acceleration of falling bodies, these factors are annoying interferences to be ignored or to be averaged away. Second, certain such effects are invisible in the ordinarily reported averaged data and so are often undetected by the experimenter.

## An Alternative Theoretical Perspective

A different issue from these technical concerns is the general approach taken to understanding psychophysical choice. The classic view is directed toward measuring thresholds

(Weber's law) and scales (Fechner's and Stevens' laws) wherein it is assumed that people abstract an attribute or element of a stimulus, such as the weight of an object, and judge it in much the same way that a balance pan abstracts the weight of an object. In this manner, the classic psychophysical approach mimics classical physics.

The alternative theoretical view offered here (see also Lockhead, 1972, 1992) is based on evidence that people cannot abstract elements and on the assumption that organisms did not evolve for this purpose. There is no a priori reason to expect people to be good balance pans or light meters or sound meters. Rather, among many other pressures, people evolved to detect and identify objects and situations in complex and ever-changing worlds. For these purposes, what must be abstracted and judged from the stimulating environment is not the intensity of a feature of an object but the object itself.

Serving as a meter to measure elements would interfere with this goal of identifying objects because, to repeat a prior example (Lockhead, 1992), the light coming from the fur of a moving tiger changes as the tiger runs in and out of the shadows, while the variously changing object must constantly be perceived as a tiger. This cannot be accomplished by abstracting the intensity of the light coming from the fur because that light changes almost randomly and thus is not informative. Rather, this is accomplished by processing some form of energy differences.

According to this view, *object constancy is fundamental to perception and attribute scaling is not fundamental*. Measuring the light intensity of the tiger's fur can only interfere with the object identification process, because that intensity changes from moment to moment; however, if what we measure is some sort of relation or difference between that energy and surrounding energies, then the tiger's appearance changes very little as it moves about. These differences or contours are essentially the same in the light as in the shadow and are the bases for perception.

The processes involved here may have analogy to calculus, wherein derivatives of functions provide information about changes and energy levels are unknown constants. Some such process must be a basis for the psychological constancies that evolved to provide essential information for object identification. These, not energy amounts, are the basic psychophysical processes that need to be understood.

## Experimental Design in Classical Studies

It is standard experimental practice in psychophysical studies to hold constant everything that is irrelevant to the task. The logical reason is that only then can one be sure that experimental results are due to an independent variable, such as light intensity, and not to confounded factors. But this practice introduces two uncontrolled problems.

The first experimental design problem is that holding everything constant except the level of the independent variable confounds that attribute or element or feature, such as light intensity, with *differences* between it and the background; when only the element intensity increases in such studies involving fixed backgrounds, the difference between it and the black (for example) background also increases, and thus the ratio between these intensities changes as well. That is, the magnitude of the element and the magnitude of the difference or ratio between it and the fixed background are perfectly correlated with one another, and so there is no way to know, in such studies, which determines the response. As noted earlier, the obverse of this also occurs: The same fixed light appears brighter on a dark background than on a light background, which means that brightness is a function of some difference between the intensity of the element and its surround, not simply intensity (e.g., Heinemann, 1955).

A related difficulty with holding all features constant except for one feature of a stimulus object is that this confounds the feature with other stimulus aspects and with the object itself. As with the background, when the classic procedure is used, there again is no way to know whether people initially process the element of the stimulus as instructed by the experimenter, process the entire stimulus object, process some difference between that element and other elements or features of the stimulus object, or something else.

I became sensitive to this problem many years ago while I was conducting a magnitude estimation study of loudness. Halfway into the experiment, I intentionally turned the oscillator dial on the sound generator, which changed the

pitch of the next tone. As a result, the participant came out of the booth to report that something was wrong with the equipment. This should not happen if people judge only intensity.

The second experimental design problem is that the traditional analysis method in psychophysical studies is to average all of the responses to each stimulus across trials. This practice confounds effects of the stimulus with effects of relations between it and other stimuli included in the experiment. One aspect of this issue is seen in an absolute judgment experiment involving three stimulus tones that differed only in intensity (Gravetter & Lockhead, 1973). Tones were randomly selected and presented one at a time for identification on each of many trials. Two of the tones were held constant throughout the study, whereas the third tone had different intensities (loudness) in different conditions. In each condition, the observers pressed one of three keys to identify the tone on each of many trials. The data of interest here involve confusion between the two tones that remained fixed across conditions; responses to these tones are more variable in conditions in which the third stimulus is more different from them. In this particular study, response variability increased linearly with the square of the stimulus range.

This result may be counterintuitive, because it means that fixed stimuli become more and more difficult to identify as other stimuli are made more different from them. One might expect fixed tones to be easier to identify when the third one is more different, because then it is less easily confused with the tones of interest. (The outcome is essentially the same whether the third tone is quieter or louder than the two fixed tones.) To explain this result, we suggested a microscope-type model in which less detail is seen when the microscope is set to low magnification (i.e., the field of view must be large to encompass the entire range) than when it is set to high magnification (the field of view is small). In this optical model, just as in the judgment data, there is a trade-off between range and precision.

This range model fits the data very well. However, like so many theories that are well fit by data, the model is wrong.

## Sequence Effects

This range model is wrong because the same range effects occur within conditions as between conditions (Lockhead & Hinson, 1986). Because range models are based on the range of items in the total set, the optical model does not predict this result; the reason is that the total set is fixed when sequence events are measured within conditions, and thus there is no range difference. That is, not only is performance poorer when the range is larger in one condition than in another condition; performance is also poorer on trials in which the difference between successive stimuli within a condition is larger. This within-trial sequence effect predicts the overall range effect, but the obverse is not true.

This within-trial effect is quite general; it occurs in absolute judgment and magnitude estimation data, whether there are many or few stimuli, with a variety of stimulus dimensions, and when the subjects are pigeons as well as people (Hinson & Lockhead, 1986; Lockhead, 1992). Apparently, judgments are based on observers' comparisons of each stimulus with previous events.

## Are Range Effects a Result of Comparing Successive Stimuli in Memory?

Range effects occur in psychophysical tasks because people judge relations between the current trial and memories of what preceded it (King & Lockhead, 1981; Treisman & Williams, 1984). When successive stimuli are perceptually similar, they are rather easily compared, and performance is relatively precise; however, when successive stimuli are more different from each other, they are more difficult to compare, and performance is more variable. In some psychological tasks, such sequence effects are called priming effects. According to this interpretation, performance is poorer (less accurate and more variable) in large range conditions than in small range conditions because the average difference between successive stimuli is larger.

Trial-to-trial sequential measures demonstrate marked assimilation and capture much, but not all, of the response variance in such judgment data (Holland & Lockhead, 1968; Huettel & Lockhead, 1999). This might not yet be explicit for several reasons, including recep-

tor noise, overall set effects that are not understood, the fact that observers are not always attentive, and sequence effects that have not been measured. Nonetheless, so much response variability is accounted for by sequential measures that it is worth reconsidering ideas other than psychophysical scaling issues that are also based on the thesis that stimulus features, elements, and attributes are processed independently. Three such ideas briefly considered here are channel capacity, local versus global processing, and statistical decision theory (SDT).

## Channel Capacity

Based importantly on work conducted by George Miller and Donald Broadbent in the 1950s, the concept of channel capacity has become part of psychological thinking. The general idea is that the capacity for processing information is limited. The specific idea is that there is a surprisingly small limit to the number of categories into which people can reliably classify univariate events.

The primary empirical impetus for the idea of a channel capacity of $7 \pm 2$ univariate category members came from absolute judgment experiments showing that once the number of response categories becomes greater than the channel capacity, further increases in the number of stimuli or in the spacing between stimuli do not produce increases in the amount of information transmitted (Miller, 1956). This capacity inference generated many theoretical debates, some of which involved independent versus dependent issues, chunking, and models of choice; all such ideas are based on the assumption that stimulus elements are what are categorized. Because the core assumption that responses are made to elements is not supported, none of these theories are appropriate, and a different interpretation is proposed here.

In classic studies of channel capacity, just as in classic psychophysical scaling studies, successive stimulus differences are confounded with the number of categories. This is because when there are more equally spaced categories (e.g., stimuli), the average trial-to-trial difference between successive categories to be judged is larger. This fact allows the suggestion that the channel limit is due to differences rather than to elements or to the number of categories involved in the situation.

Channel capacity arguments are largely based on absolute judgment experiments. To conduct such studies, experimenters determine a difference between two attribute intensities, such as two loudnesses or dB levels, according to which these intensities can be discriminated reliably. Then they add more and more stimuli with this same spacing to ensure that the discriminability between all adjacent items is approximately equal throughout the set, and—in different experimental conditions—they continue to add items until the information transmitted no longer increases. This information amount is the measure of the channel limit and is commonly less than three bits, about five or six categories. A schematic example may make the idea clear. An initial stimulus spacing is selected to ensure that the items can be discriminated along the dimension being examined, such as X X, where the distance between Xs is at least one just noticeable difference along the dimension being studied (e.g., weight). Then more weights are added to the set to determine how well they can be identified: X X X X. Subsequently, more are added in different conditions, X X X X X X X X, until capacity is reached, which is when the amount of information transmitted in an identification task no longer increases. Unfortunately, this method ensures that overall stimulus range, number of stimuli, and average difference between successive stimuli within the study all increase together and thus are confounded with one another.

One way to disentangle stimulus number from successive intensity differences is to increase the physical difference between successive stimuli in the set while holding the total number of stimuli constant, such that the spacing across conditions changes, for example, from X X X X X X X to X X X X X X X. These added separations between adjacent stimuli make them easier to discriminate, whereas the resulting increased trial-to-trial differences between successive stimuli make them more difficult to identify.

In general, as the number of stimulus members increases from two to many, and as successive stimulus differences increase from small to large, information transmission increases until a plateau, after which information transmission no longer increases. This point is where the advantage of separating adjacent

stimuli, which makes them easier to distinguish, is balanced by the disadvantage of increasing the differences between successive stimuli, which makes precision more variable. It appears that channel capacity is largely the result of these two factors, pairwise discriminability and successive item differences, and is not associated with the number of categories.

This conclusion that number of categories or stimuli is not the limiting factor for performance is supported by the three-stimulus study described earlier in which performance became increasingly poor as one item was made increasingly more different from the others. Because this occurs well below any proposed number limit, channel capacity is not associated with numbers of categories, as is commonly assumed, but instead is associated with comparative judgments that are noisier when the difference between successive trials is larger.

## Local and Global Psychophysics

Many psychophysicists study stimulus confusion. This category of research is called *local psychophysics* "because the focus is on stimulus changes that are small enough to cause confusions among stimuli" (Luce & Krumhansl, 1988, p. 39). Other psychophysicists deal primarily with stimulus differences that vary over such a large physical range "that there is no chance whatever of confusions between the extreme signals in the range" (Luce & Krumhansl, 1988, p. 39). These topics are called *global psychophysics*.

One might expect local and global psychophysics to be two parts of a common issue, but this has not been found to be the case. Rather, relating findings across methods is so problematic that creating a bridge between *local* and *global* psychophysics has been called "the oldest theoretical problem in psychophysics[;] it remains unresolved" (Luce & Krumhansl, 1988, p. 39). However, it appears possible to reconcile this issue and place these two areas into a single framework by using the same argument here as was used earlier for channel capacity.

## Confounding Range and Number

The absolute judgment tasks that provided the concept of channel capacity involved the use of large stimulus ranges and thus were classified

under *global psychophysics*, but the stimuli used in those studies were selected in pairwise tasks classified under *local psychophysics*. This can be seen as follows: In selecting stimuli for such experiments, one measures discriminability between adjacent intensities using a local technique, such as two-alternative forced choice. This comparative judgment task is used to ensure that adjacent stimuli can be discriminated. It is traditional to use "same" and "different" as the responses to each stimulus in these studies, but "1" and "2" would work as well and might better reveal the similarities between these classes of studies. Here stimulus range is small, and the range between successive stimuli is also small.

Ideally, successive pairs of stimuli are selected and measured for discriminability and then are strung together to create a set of many stimuli to be examined with global measures such as absolute judgment or magnitude estimation. In practice, however, once one stimulus separation is selected by means of a local method—perhaps the choice is 2 dB for a loudness study—all adjacent stimuli in the set are separated by 2 dB.

On some trials in the global task successive stimuli are very similar (small range), and on other trials they are very dissimilar (large range). The thesis here is that this addition of different ranges to the initial two-alternative comparative judgment task produces the apparent (only) discrepancy between local and global psychophysics. If this is approximately correct, then:

1. The idea of a channel capacity for numbers of elements in absolute judgment tasks should be reinterpreted as a capacity for the precision of comparative judgments.

2. Local and global psychophysics differ only in the stimulus ranges studied. People make comparative judgments in both situations.

3. There is only one psychophysics, and the local–global distinction should be replaced with processes involved when people compare their perception of the current event with memories of what preceded it.

## Statistical Decision Theory: Sensory Noise or Criterion Noise?

SDT provides valuable information about discriminability between events and about cri-

terion placements used for decision. SDT might provide even more information if one feature of the theory were used differently: Rather than treat variability in the data as sensory noise, as is commonly done, it could be treated as an effect of memory that shifts the criterion between trials.

Most commonly, SDT is used with a two-alternative forced-choice procedure in which two levels of the stimulus (noise only and signal plus noise) and two levels of the response ("yes," it was a signal or "no," it was not a signal) are cast into a $2 \times 2$ matrix. On trials in which a signal occurs the observer says either yes or no, and on trials in which there is no signal the observer again says either yes or no. There are two degrees of freedom in the matrix, traditionally expressed as the hit rate, which is the proportion of times a presented signal was correctly called a signal, and the false alarm rate, which is the proportion of times no signal was presented but the observer reported a signal. Both are needed to measure discriminability ($d'$) and criterion bias ($\beta$). Accordingly, there is no information remaining to determine whether the noise seen in SDT studies should be ascribed to the input or to placements of the criterion, and there is no way to reduce further the variability in the data in terms of discriminability.

Within the theory, variability in the data is due to the input (e.g., noisy signal or receptors), the output (e.g., faulty memory or attention), or some combination thereof. Green and Swets showed in 1966 that it is formally equivalent to assign noise to the output and to the input, and the historical solution has arbitrarily assigned it to the input. Some have suggested, but have not demonstrated, that the criterion might be the primary source of performance variability, and information from outside the theory suggests they are correct.

Outside SDT, additional information comes from event sequences, which are structured, and this structure allows discrimination between the two alternatives, input or output. The reasoning is direct. In every judgment study in which sequence effects have been measured, responses depend in part on the previous trial(s); the same stimulus is assigned different responses on different trials depending on what preceded it. This dependency must be the source of at least some of the response variability that ordinarily is

treated as noise in SDT studies. However, noise is an incorrect designation, because this variability is structured; it is not random, and so it is not noise. To the extent that this variability can be measured, it can be removed from SDT data to provide improved measures of discriminability and improved predictions of what observers report in successive judgment tasks. Furthermore, if the source of this structure can be determined, then the question of whether variability ought to be allocated to the input or to the output in STD studies might be answered.

To my knowledge, Morris Holland (1969) was the first person to show that response variability in psychophysical tasks is partially due to variability in the criterion. Holland's work was not published beyond his dissertation, because reviewers argued that his studies did not capture all of the variance in the data; however, he did demonstrate that, and how, the position of the criterion varies across intertrial durations and stimulus ranges, and I recognize him here for beginning this argument. That thesis led in part to the three-stimulus study conducted by Gravetter and Lockhead (1973), described earlier, showing that overall stimulus range is correlated with the variability of the criterion and that $d'$ decreases as the range of other stimuli in the set increases. The fact that discriminability between events becomes increasingly less competent as another stimulus is made more different means that the criterion for classifying the fixed objects is more variable when other stimuli in the set vary over a wider physical range. That is, the criterion is more variable.

Accordingly, much of what is usually called noise in judgment data is not a result of random activity but represents structure due to memory factors that, in turn, result in shifting criterion placements. This supports the conclusion here that some or all of the variability in SDT studies should not be labeled as noise at the input but should instead be labeled as a memory effect on the position of the criterion.

## A Demonstration of a Memory Effect in Judgment

If the preceding arguments are generally correct, then people do not identify a stimulus as presented but do identify its relation to memories of other events, many of which depend on memories of preceding events. To examine this

more directly, I asked four people to make 200 magnitude estimations of the loudness of six different tones, after which I unexpectedly asked them to estimate how many different tones were being used in the study.

Although there were only 6 tones, the guesses ranged from 62 to 100 different tones. Even on the next day, when the study and question were repeated with the same observers, the guesses ranged from 38 to 82 tones. Nothing like this should occur if what observers do in such psychophysical tasks is abstract and identify intensities. But such overestimations should occur if each tone is heard as a difference between it and the memory of what preceded it and if that memory is biased by what preceded it. Apparently, each of the $N$ tones is compared with $N^3$, or even more, memories that preceded the tone. This explains why many more than $N$ stimuli are perceived and remembered when, for example, only 6 are used, and it is a strong case for the argument that people judge relations, not absolutes, in psychophysical judgment tasks.

## References

Braida, L. D., & Durlach, N. I. (1988). Peripheral and central factors in intensity perception. In G. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Auditory function* (pp. 559–583). New York: Wiley.

Fechner, G. (1966). *Elemente der Psychophysik* [Elements of psychophysics] (Vol. 1, H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing- and test-field luminances. *Journal of Experimental Psychology, 50,* 89–96.

Hinson, J., & Lockhead, G. R. (1986). Range effects in successive discrimination procedures. *Journal of Experimental Psychology: Animal Behavior Processes*, 12, 270–276.

Holland, M. K. (1969). Channel capacity and sequential effects: The influence of the immediate stimulus history on recognition performance (Doctoral dissertation, Duke University). *Dissertation Abstracts International, 30,* 1923B.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics, 3,* 409–414.

Huettel, S. A., & Lockhead, G. R. (1999). Range effects of an irrelevant dimension on classification. *Perception & Psychophysics, 61,* 1624–1646.

King, M., & Lockhead, G. R. (1981). On memory effects in magnitude estimation experiments. *Perception & Psychophysics, 30,* 599–603.

Laming, D. (1997). *The measurement of sensation.* Oxford, England: Oxford University Press.

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review, 79,* 410–419.

Lockhead, G. R. (1992). Psychophysical scaling: Judgments of attributes or objects? *Behavioral & Brain Sciences, 15,* 543–559.

Lockhead, G. R., & Hinson, J. (1986). Range and sequence effects in judgment. *Perception & Psychophysics, 40,* 53–61.

Lockhead, G. R., & King, M. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 461–473.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics, 20,* 49–54.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling and psychophysics. In R. Atkinson, R. Herrnstein, G. Lindzey, & R. Luce (Eds.), *Stevens' handbook of experimental psychology* (2nd ed., pp. 3–74). New York: Wiley.

Luce, R. D., & Suppes, P. (2002). Representational measurement theory. In J. Wixted & H. Pashler (Eds.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 1–42). New York: Wiley.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science, 133,* 80–86.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91,* 68–111.

Weber, E. H. (1965). On the sense of touch and common sensibility. In B. Haupt (Trans.) and R. J. Herrnstein & E. G. Boring (Eds.), *A sourcebook in the history of psychology.* Cambridge, MA: Harvard University Press. (Original work published 1846)