MEASURING PERCEIVED MAGNITUDE: IT'S LEGAL BUT IS IT LAWFUL?

Robert Teghtsoonian
Smith College
Northampton, MA 01063, USA
rtex@smith.edu

Abstract

The study of perceived magnitude has been a central issue in psychophysics since Fechner proposed that it is proportional to the logarithm of stimulus strength. There is now a large body of evidence showing that relation to follow a power law, and that it holds for a wide variety of perceptual continua. While the methods developed for the measurement of perceived magnitude have grown in number and variety, and have been successfully applied to a wide variety of real-world problems, critics have raised questions about the power law, and whether it or any other general rule can be shown to be generally applicable. This paper reviews such objections and argues that alleged deficiencies can be resolved, and that for the present, the power law is too valuable, as a summary statement encompassing a large body of experimental findings, to be abandoned.

We observe October 22 as Fechner Day in honor of Gustav-Theodor Fechner, the founder of the field of psychophysics, the oldest (and perhaps the most refined) subject of study in psychological science. A complete history remains to be written, but a more modest goal may be to consider how one topic within that field has evolved since the founding of the International Society for Psychophysics 25 years ago. I've chosen the area I know best, the study of perceived magnitude, the problem of how to describe what we experience when presented with signals that may vary in strength. In particular I'll consider how perceived magnitude has been measured, and whether those measurements have revealed a fundamental relation between perceived magnitudes and the stimuli that produce them.

During the past 25 years the market for psychophysical expertise has flourished. The need for methods to measure perceived magnitudes is now widely recognized. What began as a search for ways to measure the properties of the sensory systems is now an important approach to many other topics. Cross-modal matching (CMM) has been used in studies of eating behavior to measure perceived satiety (M. Teghtsoonian, 1981). The degree of fear experienced by subjects as a function of the distance from the feared object has been reported (R. Teghtsoonian & Frost, 1982). And there is now a scale of clutter intended for use with pathological hoarders (Frost et al. 2007). The hedonics of food and beverages (Moskowitz, 1981) and the impact of noise sources in the environment (Berglund, 1991) have been explored with psychophysical scaling techniques. Borg (1998) has devised scaling procedures that are now standard in two quite separate fields, exercise science, and the medical study of pain. Nor is the popularity of psychophysical measurement limited to just the psychological and medical sciences. In recent years economists have discovered that they are really students of human behavior and have rushed to find ways to make relevant measurements. Fechner would be dazzled by the degree to which his basic idea, that perceived magnitude could be measured, has become almost universally accepted and now flourishes in so many fields of study.

In addition to new topics there is a continuing growth in the number of methods in use. Older methods such as magnitude estimation (ME), magnitude production (MP), category judgments (CJ), reaction time, conjoint measurement, and Berglund's master scale, all continue in use. More recently Bartoshuk et al. (2004) have described a generalized form of a Labeled Magnitude Scale calling for the use of "adjective/adverb intensity descriptors." West, Ward & Khosla (2000) advocate what they call constrained scaling, a method in which subjects are trained with feedback to produce magnitude estimates that are power functions of signal level with a specified exponent. Visual Analog Scales (see e.g., Torrance et al., 2001) define end-points, sometimes in the form of a straight line, and ask the observer to represent the perceived magnitude of a stimulus by marking a point on that line. Susini, McAdams, & Smith (2002) describe a technique (used by them and others) for judging the perceived magnitude of a stimulus that is continuously changing in intensity: Observers are asked to make a continuous viewable tracing representing their continuously changing perceptions. Schneider & Parker (2009) have proposed a non-metric scaling procedure.

On the face of it, more methods may be seen as an unalloyed good. Measurement lies at the heart of psychophysics, so adding to the list of measurement techniques could be seen as a sign of growing strength and versatility. But some aspects of this growth are troubling. There may be features of new methods that simply reflect the whims of the experimenters who use them. Scale end-points, verbal labels, and instructions to the subjects can and do vary widely, often with little or no rationale. Journal editors could be more aggressive in pressing for details. Is this a new method, and if so what evidence is there for its validity? Is it a variant of an older method, and if so what justification is there for the new features? Weak answers to such questions should lead to summary rejection, and as a result we would all be spared a lot of methodological clutter.

But there's a more basic issue that emerges in considering the wide variety of measurement techniques in current use, and that is the relation to older methods and the results obtained with them. Some of the new techniques have been designed to replicate the findings of earlier methods while eliminating what were seen as disadvantages. The constrained scales developed by West, Ward & Khosla (2000) were validated by comparison with conventional magnitude estimation (ME) scales. Much the same tactic was involved in some of the Borg category scales: The number of categories and their placement on a continuum was at least in part managed to achieve agreement with results obtained with magnitude estimation. In both cases the rationale for the new features was the belief that they would make it easier to interpret differences among individuals or across test situations. Whether those beliefs are justified can be debated, but there is little doubt about the criterion for evaluation: The new method should yield results consistent with earlier findings obtained with magnitude estimation.

In contrast to such studies, others raise a fundamental question about the merits of Stevens's psychophysical law. Using some variant of traditional scaling techniques, these studies report either a failure to replicate the power law, or show its parameters to be so dependent on contextual factors as to make the claimed normative values meaningless. A few examples can be provided both as illustrations of this approach and to assess their implications for any kind of general law relating perceived magnitudes to signal amplitudes.

a) Range effects.

Poulton (1967), among others, has proposed that the results of ME scaling studies are so sensitive to the range of stimulus intensities presented, that no single result can be regarded as definitive. But the choice of range should not be a matter of the investigator's whim. Why not pick some defined proportion of the dynamic range for all perceptual continua? R.

Teghtsoonian & M. Teghtsoonian (1997) showed that this could be done by asking individual observers to choose the limits of the test range for five different continua, using the same criteria in each case. Those observer-defined ranges were used both to predict their performance on a scaling task, and to generate estimated exponents in close agreement with standard values in the literature. We concluded that range sensitivity is not a fatal flaw in magnitude estimation scaling (R. Teghtsoonian, 2004).

b) Stimulus Density.

Luce & Mo (1965) obtained loudness and heaviness judgments from individual observers, and used many closely spaced stimulus levels and many judgments per stimulus level in order to obtain a more reliable and finer-grained view of the matching functions for individual observers. For advocates of a simple power law the results did not yield a pretty picture: For both continua, individual functions followed oddly sinuous paths on log-log coordinates, appearing to accelerate and decelerate in unpredictable ways. But later research (R. Teghtsoonian, M. Teghtsoonian, & Baird, 1995) showed that such sinuosity was in fact systematically predictable, based on the assumption that judgments reflected, not only stimulus strengths but also a bias exhibited in favor of numbers that are simple powers of 5 and 10. Again, a procedural variation is shown to reveal strange departures from a power law, but these particular departures turn out to be easy to model, and, potentially, to eliminate.

c) Scaling multidimensional stimuli.

Perceptual attributes are often dependent on more than one independent variable. For example, the judged heaviness of a lifted object is governed by both its size and its weight. Anderson (2004) cited an early study showing ratings of heaviness as a function of cylinder height (with weight as a parameter) to be linear. Here the argument is not that there is no general law, but instead that the power law is the wrong one. Yet there is no mention of two studies (J. Stevens & Rubin, 1970; Cross & Rotkin, 1975) using ME and much larger ranges of both stimulus dimensions. Quite beautifully, what they found is a family of converging functions relating judged heaviness to weight with size as the parameter: For very large weights, size matters not at all. When the weight is very large, and approaches the capacity of the observer to lift it, all sizes feel equally heavy. Once again, seemingly disconfirming data are negated by compelling evidence supporting a power law. (In passing one should note that trying to estimate the form of a function with a stimulus range covering a ratio of only 3 to 1 (as did Anderson) is unwise; Cross and Rotkin used a range of weights spanning a ratio of about 100 to 1; Stevens & Rubin, about 250 to 1).

d) Direction of change and direct judgment of perceived change.

In recent years there has been growing interest in the experience of loudness when the sound level is changing in a continuous way (e.g., Canévet & Scharf, 1990; Neuhoff, 1998; R. Teghtsoonian, M. Teghtsoonian, & Canévet, 2005). Neuhoff (1998) argued that judged change in loudness is greater when the direction is upward even though the change in level is the same as when the direction is down, whereas a simple power law for loudness predicts no difference. But it may be that judgments of change are not as simple as they seem. When listeners were asked to judge the amount of change in a sweep from a moderate level to a higher end-point, the results were only weakly related to the actual size of the physical change. But when judged change was plotted against the strength of the end level, it followed virtually the same form as would have been expected if the listeners had been asked to judge directly the loudness of the end level (R. Teghtsoonian, M. Teghtsoonian, & Canévet, 2005). One must be concerned about the merits of a dependent variable that can be so insensitive to the nominal antecedent. Judgments about the interval separating stimulus pairs has achieved some popularity (e.g., Schneider & Parker, 2009) but we think that caution is needed in evaluating such data, and we remain skeptical about their implications for the validity of a power law for perceived magnitudes.

e) Sequence effects.

Ideally, an observer presented with a sequence of varied stimulus intensities would be able to make each judgment independently of the preceding trials. Some individuals seem able to do this, and for those who can't, at least one technique has been developed to make bias-free judgment possible (M. Teghtsoonian, R. Teghtsoonian, & DeCarlo, 2006). Yet in many studies, it's clear that sequence effects do occur, and as a result it becomes difficult to assess the degree to which any given judgment is due to the stimulus level, or to preceding stimuli and previous judgments. But there is a solution: DeCarlo (1994) has developed a mathematical model and a theory of sequence effects that makes it possible to separate these factors, and provide bias-free estimates of the effects of stimulus strength on judgments of perceived magnitude.

Conclusions

The work of S.S. Stevens (1975) included a large body of data showing a power law to be a better description of the relation between perceived magnitude and stimulus strength than was Fechner's logarithmic law. And there continues to be a stream of research, both empirical and theoretical, deriving from Stevens's work. Reports by Narens (1996), Luce (2002), Ellermeier & Faulhammer (2000), Link (1994), Baird (1997), Norwich (1993), and R. Teghtsoonian & M. Teghtsoonian (1997) among many others, agree (more or less) with the fundamental relation described by the power law.

Nevertheless, some critics have reported evidence challenging the entire notion of a consistent relation between judged and physical magnitudes. Lockhead (1992) cited a large body of such findings and concluded that "any underlying…psychophysical scale can only appear in the data as a will-o'-the-wisp." A similar conclusion was reached by Schneider & Parker (2002): They argue that, context being decisive in the result of any attempt to measure perceived magnitude, psychophysics should abandon its search for basic laws relating stimulus intensity to perceived magnitudes, and should devote itself instead to the study of the many ways in which context affects judgment. How should we respond to this diversity of opinion?

There are several points that need be considered: First, the amount of evidence in support of a power function for perceived magnitude is very large, and cannot be sensibly abandoned on the basis of isolated reports of disconfirming evidence. It seems to me inescapably clear that, for most intensive continua, the relation between perceived magnitude and stimulus level is non-linear, and that, at least to a first approximation, the non-linearity takes the form of a power function. Will there be departures from that simple rule? The answer will almost certainly be yes, as for example in the account of the form of the loudness function provided by Florentine (2006). But it remains true that if we want to make the best prediction about the perceived change due to a fixed ratio change in intensity at different points across the dynamic range for almost any perceptual continuum, we will predict a constant ratio change. Given the amount of evidence supporting such a prediction, it seems perverse to argue that such predictions have no merit, and that there is no general principle underlying them.

Second, although the bulk of the evidence supporting the power law was obtained with the method of magnitude estimation, there is now almost universal recognition that it is only a special case of cross-modal matching. Indeed, ME data can be used in exactly the same way as any other matching data. When taste intensity was judged by ME and by the loudness of a matching tone, the ratio of the exponents of the resulting power functions correctly predicted the exponent for a match of number to sound level (Moskowitz,1971), a finding that has been confirmed at the level of individual observers (Daning, 1983). If we know the exponents for the power function between continuum A and continuum C, and for B and C, we can predict the exponent for the function relating A to B. The existence of this coherent network of matching data seems to me one of the great achievements in the field of psychophysics and should be recognized as such. I have tried to spell out in detail the meaning of this network and have proposed an integrative system that encompasses the data for virtually all intensive continua (e.g., R. Teghtsoonian, 1974; M. Teghtsoonian & R. Teghtsoonian, 2003). There has yet to be a compelling critique of either the empirical evidence showing the consistency of findings using ME and CMM, nor of the integrative system that I've proposed. Until such a critique appears, both merit continued attention.

Third, I would add a cautionary note about the pool from which we draw the participants in our research. Whether observers are believed to respond to ratios or differences or simple rank order, and whether their judgments are thought to reflect subjective ratios, subjective differences, or subjective ranks, may in the end be, not so much a matter of theory as it is an empirical question, about who the observers are and what stage of intellectual development they have achieved. Having for many years drawn my observers from a pool of undergraduates in a selective liberal arts college, I can say unequivocally that many, possibly more than half, do not understand the concept of ratios. And I can imagine many subject pools in which that proportion may be even higher. Perhaps the greatest risk we take in using numerical magnitude estimation is the ambiguity we face in judging a subject's capacity to use numbers. We need to take seriously the possibility that the form that our data takes may depend on the particular group of observers who create those results. Magnitude estimation is quick and easy, but the gold standard should be cross-modal matching.

Finally, in thinking how best to respond to those who report data that conflict with predictions of the power law, I can think of nothing more relevant and more foresighted than the final paragraph of Stevens's *Psychophysics*:

"Although some theorists seem impelled to reject the idea that there exists any binding and lawful relation between stimulus and response, other psychophysicists seem prepared to acknowledge that Fechner posed a valid question when he inquired into the relation between sensation and the intensity of the physical configuration that gives rise to it. Those who believe that no psychophysical laws are possible and who reject the nomothetic imperative find their belief nurtured by the ease with which an experiment can be made to yield contrary and discordant results. And those who believe that sensation obeys discoverable laws find their faith reinforced by the results of numerous experiments that accord with the principle that equal stimulus ratios produce equal sensation ratios. That simple invariance stands as the psychophysical law. The measurement procedures that have been devised for its quantification suffer from an overlay of contextual factors that bias the outcome of each particular experiment. But those biasing factors seem also to obey rules that can be discovered. And those rules in turn enable us to discern better the underlying simplicities…"

References

Anderson, N.H. (2004). Psychophysical law and psychological law. In A.M. Oliveira, M. Texeira, G.F. Borges & M.J. Ferro (Eds.) *Proceedings of the Twentieth Annual Meeting of the International Society for Psychophysics* (pp. 32-43). Coimbra, Portugal: ISP.

Baird, J.C. (1997). *Sensation and Judgment: Complementarity Theory of Psychophysics.* Mahwah, New Jersey: Erlbaum.

Bartoshuk, L.M., Duffy, V.B., Green, B.G., Hoffman, H.J., Ko, C.-W., Lucchina, L.A., Marks, L.E., Snyder, D.J., & Weiffenbach, J.M. (2004). Valid across-group comparisons with labeled scales: the gLMS versus magnitude matching. *Physiology & Behavior*, 82, 109-114.

Berglund, M.B. (1991). Quality assurance in environmental psychophysics. In S.J. Bolanowski & G.A. Gescheider (Eds.) *Ratio Scaling of Psychological Magnitude* (pp. 140-162). Hillsdale, New Jersey: Erlbaum.

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales.* Champaign, IL: Human Kinetics.

Canévet, G., & Scharf, B. (1990). The loudness of sounds that increase and decrease continuously in level. *Journal of the Acoustical Society of America,* 88, 2136-2142.

Cross, D.V., & Rotkin, L. (1975). The relation between size and apparent heaviness. *Perception & Psychophysics*, 18, 79-87.

Daning, R. (1983). Intraindividual consistencies in cross-modal matching across several continua. *Perception & Psychophysics*, 33, 516-522.

DeCarlo, L. T. (1994). A dynamic theory of proportional judgment: Context and judgment of length, heaviness, and roughness. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 372-381.

Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens' ratio-scaling approach: I. Loudness production. *Perception & Psychophysics,* 62, 1505-1511.

Florentine, M., & Epstein, M. (2006). To honor Stevens and repeal his law (for the auditory system). In D. E. Kornbrot, R. M. Msetfi, & A. W. MacRae (Eds.) *Fechner Day 2006: Proceedings of the 22nd Annual Meeting of the International Society for Psychophysics* (pp. 37-41). St. Albans, UK: The International Society for Psychphysics.

Frost, R. O., Steketee, G., Tolin, D. F., & Renaud, S. (2007). Development and validation of the Clutter Image Rating. *Journal of Psychopathology and Behavioral Assessment*, 30, 193-203.

Link, S. W. (1994). *The Wave Theory of Difference and Similarity*. Hillsdale, NJ: Erlbaum.

Lockhead, G. R. (1992). Psychophysical scaling: Judgments of attributes or objects? *Behavioral and Brain Sciences,* 15, 543-601.

Luce, R. D., & Mo, S. (1965). Magnitude estimation of heaviness and loudness by individual subjects. *British Journal of Mathematical and Statistical Psychology*, 18, 159-174.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, 109, 520-532.

Moskowitz, H. R. (1971). Intensity scales for pure tastes and taste mixtures. *Perception & Psychophysics*, **9**, 51-56.

Moskowitz, H.R. (1981). Sensory intensity Vs hedonic functions: Classical psychophysical approaches. *Journal of Food Quality*, 1981, 5, 109-138.

Narens, L. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology*, 40, 109-129.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, 395, 123-124.

Norwich, K. H. (1993). *Information, Sensation, and Perception*. San Diego, CA: Academic Press,

Poulton, E. C. (1967). Population norms of top sensory magnitudes and S. S. Stevens' exponents. *Perception & Psychophysics*, 2, 312-316.

Schneider, B. A., & Parker, S. (2002). Psychophysics in the 21st century: Abandoning the quest for sensory scales. In J. A. da Silva, E. H. Matsushima, & N. P. Ribeiro-Filho (Eds.) *Annual Meeting of the International Society for Psychophysics*. Rio de Janeiro, Brazil: International Society for Psychophysics.

Schneider, B. A., & Parker, S. (2009). Human Methods: Psychophysics. In Squire, L. R. (Ed.), *Encyclopedia of Neuroscience Volume 5*, (pp. 19-27). Oxford: Academic Press.

Stevens, J. C., & Rubin, L.L. (1970). Psychological scales of apparent heaviness and the size weight illusion. *Perception & Psychophysics*, 8, 225-230.

Stevens, S. S. (1975). Psychophysics: Introduction to its perceptual, neural, and social prospects. New York: Wiley.

Susini, P., McAdams, S., & Smith, B. K. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acustica United with Acustica*, 88, 536-548.

Teghtsoonian, M., Becker, E., & Edelman, B. (1981). A psychophysical analysis of perceived satiety: Its relation to consumatory behavior and degree of overweight. *Appetite: Journal for Intake Research*, 2, 217-229.

Teghtsoonian, M., & Teghtsoonian, R. (2003). Putting context effects into context. In B. Berglund & E. Borg (Eds.), *Proceedings of the Nineteenth Annual Meeting of the International Society for Psychophysics* (pp. 309-314). Stockholm: ISP.

Teghtsoonian, M., Teghtsoonian, R., & DeCarlo, L. T. (2008). The influence of trial-to-trial recalibration on sequential effects in cross-modality matching. *Psychological Research*, 72, 115-122.

Teghtsoonian, R. (1974). On facts and theories in psychophysics: Does Ekman's law exist? In H. R. Moskowitz et al. (Eds.), *Sensation and Measurement* (pp. 167-176). Dordrecht, Holland: D. Reidel.

Teghtsoonian, R. (2004). Range effects: How many and how important? In A.M. Oliveira, M. Texeira, G.F. Borges & M.J. Ferro (Eds.) *Proceedings of the Twentieth Annual Meeting of the International Society for Psychophysics* (pp. 190-195). Coimbra, Portugal: ISP.

Teghtsoonian, R. & Teghtsoonian, M. (1997). Range of acceptable stimulus intensities: An estimator of dynamic range for intensive perceptual continua. *Perception & Psychophysics*, 59, 721-728.

Teghtsoonian, R. & Frost, R. O. (1982). The effects of viewing distance on fear of snakes. *Journal of Behavior Therapy and Experimental Psychiatry*, 13, 181-190.

Teghtsoonian, R., Teghtsoonian, M., & Baird, J. C. (1995). On the nature and meaning of sinuosity in magnitude-estimation functions. *Psychological Research*, 57, 63-69.

Teghtsoonian, R., Teghtsoonian, M., & Canévet, G. (2005). Sweep-induced acceleration in loudness change and the “bias for rising intensities.” *Perception & Psychophysics*, 67, 699-712.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual analog scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making*, 21, 329-334.

West, R. L., Ward, L. M., & Khosla, R. (2000). Constrained scaling: The effect of learned psychophysical scales on idiosyncratic response bias. *Perception & Psychophysics*, 62, 137-151.