Computer Augmented Psychophysical Scaling

Robert L. West (robert_west@carleton.ca)
Department of Psychology, Department of Cognitive Science, Carleton University, Ottawa, Canada

Ronald L. Boring (rlboring@ccs.carleton.ca)
Department of Cognitive Science, Carleton University, Ottawa, Canada

Stephen Moore (srmoore@chat.carleton.ca)
Department of Cognitive Science, Carleton University, Ottawa, Canada

Abstract

In this paper we present a methodology for improving the reliability of observers in magnitude estimation tasks by using the computer to augment the cognitive components of the task.

Psychophysical scaling is the study of how to accurately measure perception. More specifically, the goal is to find methodologies that allow people to accurately communicate the magnitudes of specific dimensions of conscious experience, such as brightness, loudness, temperature, and heaviness. Psychophysical scaling can also be used for measuring the magnitude of subjective experiences such as level of happiness (e.g., West & Ward, 1988). The goal of psychophysical scaling is to find the mathematical functions that map the magnitudes of external stimulus dimensions to the conscious perception of magnitude. This enterprise is extremely useful for both scientific and applied research.

Numerous different scaling techniques exist. However, our focus is on magnitude estimation, which is one of the most commonly used psychophysical methods. Magnitude estimation (ME) was invented by Stevens (1956) and involves exposing subjects to a set of stimuli and asking them to match the magnitude of a particular dimension of each stimulus to the magnitude of a number. This is repeated for multiple trials to provide multiple responses for each stimulus value. To avoid the influence of outliers, the median or the geometric mean of the responses for each stimulus value is calculated. Numerous studies have shown that plotting these values against the stimulus values produces functions that are closely approximated by power functions. This is known as, the Power Law, or, Stevens' Law.

The form of the power law is,

$$R=KS^B,$$

where R is the observer's response, S is the stimulus magnitude, B is the exponent value, and K is a constant. Logging both sides of the equation produces,

$$\mathrm{Log}(R)=B\cdot \mathrm{Log}(S)+\mathrm{Log}(K),$$

which is a straight line with B estimated by the slope and K by the intercept. The exponent, B, can be interpreted as a metric for stimulus compression. This reflects the fact that people use a power function or something closely approximating a power function to compress stimuli, just as audio and video files can be compressed to save on bandwidth. In fact, audio and video compression go unnoticed to the extent that the compression function maps onto the human compression function for the same stimuli. Generally speaking, in ME the goal is to put as few restrictions on the observer's choice of numbers as possible. Often free ME (e.g., see Zwislocki & Goodman, 1980) is used, in which observers are instructed to match the perceived magnitude of the stimulus to whatever number seems most natural. This is quite different from the common psychological practice of imposing scales on people. The reasons for this are both theoretical and practical. From a mathematical standpoint, if any two stimuli are set equal to any two responses then you have determined what the exponent value must be. Thus, if an observer uses the lowest value on a scale to match the lowest perceived magnitude and the highest value to match the highest perceived magnitude, the power function exponent has been fixed. To get around this one could assign a value to a middle value on the scale and not impose a top end or bottom end, but this has been shown to produce confusion and poor results (Stevens, 1975). However, the fact that peoples' backgrounds cause them to use different ranges of numbers in their responses is not a problem as these differences are captured by the K constant (since response range is usually not of interest, K values are usually not reported).

ME can be considered a special case of cross modal matching (CMM). In cross modal matching, the observer adjusts the magnitude of one stimulus dimension to match the magnitude of another stimulus dimension (e.g., adjusting the brightness of a light to match the loudness of a tone). Like ME, CMM results also produce power functions. Furthermore, ME and CMM results are consistent in that they can be used to predict each other (e.g., the ME exponents for brightness and loudness can be used to predict the exponent relating brightness and loudness in a CMM experiment). Also, both the power functions and the specific exponent values found through ME are consistent with ratio scaling experiments, in which magnitude scales are derived by asking observers to set or report ratios between stimuli. These approaches to scaling are known as direct scaling techniques (Stevens, 1971).

Problems

ME forms the foundation for a potentially accurate and consistent way of measuring perceived magnitude. However, ME, as well as the other methods with which it is consistent, have been found to be limited in terms of accuracy. Although a considerable amount of evidence indicates that subjects do obey the power law (see Stevens, 1975; and Bolanowski & Gescheider, 1985 for reviews), the specific exponent values that Stevens found could not be reliably replicated with the level of accuracy one would expect for measuring sensory processes in normal, healthy individuals. Exponent values vary considerably across individuals in the same experiment (e.g., Algom & Marks, 1984; Luce & Mo, 1965; Marks & J. C. Stevens, 1965; Rule & Markley, 1971; Wanschura & Dawson, 1974; Logue, 1976) and can also vary across time within individuals (Logue, 1976; Marks, 1991; Teghtsoonian & Teghtsoonian, 1983). Stevens also found strong individual differences, which he attributed to various response biases. Stevens' solution was to treat response bias as a random factor and to average across individuals to get the true exponent value (Stevens, 1971). However, Marks (1974) reviewed the literature and found that in addition to individual differences, the average value of the exponent varies significantly across ME experiments done in different labs. These results suggest that the distribution of individual response biases differs from lab to lab, indicating that they cannot be treated as random. Indeed, it is well known that some labs get systematically higher or lower exponent values than others, suggesting that response bias can be influenced by minor procedural differences.

In addition to limitations on accuracy, ME results are not consistent with partition scaling (also called interval scaling) results for prothetic continua, although they are consistent for metathetic continua (according to Stevens, metathetic continua are more qualitative in nature, e.g., pitch or hue; while prothetic continua are more quantitative in nature, e.g., loudness or brightness; see Stevens, 1971 for a more detailed discussion). Partition scaling includes a variety of techniques that require observers to partition the stimulus continuum. Category scaling (e.g., 1 to 5 scales; 1 to 7 scales; scales partitioned by word labels such as good, bad, very bad) is a form of partition scaling, and is by far the most commonly used scaling technique. The problem is that partitioning techniques tend to produce power functions with lower exponents than direct scaling techniques (Stevens, 1971). Stevens' argument for accepting the results of direct scaling techniques rather than partition scaling techniques was that partition scaling is less direct because it requires the extra step of partitioning the stimulus range, and that the discrepancy can be attributed to biases introduced by the partitioning task (see Stevens, 1971). However, like direct scaling, partition scaling also produces excessive variability (Marks, 1974).

Because of these problems, psychophysical scaling still has issues concerning reliability and validity. In terms of the power law, the validity problem can be stated as the problem of which, if any, method will produce the "true" exponent. The reliability problem is that we do not have a methodology that we can use to make reliable statements about individual differences or inter-lab differences in exponent values. In our opinion, the reliability problem needs to be solved before tackling the validity problem. Our work attempts to address this. The reliability problem can be broken down into a theoretical and a practical problem. The theoretical problem is that if bias differs from individual to individual and within individuals across time, we cannot get reliable measurements without being able to somehow predict or control the bias. The practical problem is that even if we solve the theoretical problem, to be useful we need a system that does not require huge numbers of responses from individuals who have limited amounts of time and limited attention spans. We have focused our efforts on the reliability issue and attempted to solve both of these problems by cognitively augmenting our human observers through the use of computerized support.

Bias

The process of magnitude matching can be represented in the following way (Marks, 1991),

$$M(S) = R$$

where S is the stimulus magnitude, R is the response magnitude, and *M* is the function relating them. The *M* function can then be decomposed into an initial, perceptually based function, P, that is the same (or highly similar) across healthy, normal individuals; followed by a function, C, representing cognitively imposed constraints that account for the excessive variability:

$$M(S) = C(P(S))$$

Since most psychophysicists study perception, the emphasis has been on getting rid of C so as to reveal P. Considerable effort has been expended in this enterprise. Approaches taken include trying to identify the sources of C to avoid or control for them (see Poulton, 1989 for a review); trying to minimize C by encouraging observers to respond naturally, without thinking about it too much (e.g., Stevens, 1975; Zwislocki & Goodman, 1980); trying to measure C and then partial it out (e.g., Berglund, 1991); trying to stabilize C across scaling tasks to get rid of intra-observer variability (e.g., J. C. Stevens & Marks, 1980); and avoiding C by developing methods that allow the scale to be derived from judgments of "greater than" or "less than" for paired stimuli sets (e.g., Schneider, 1980, 1988). However, success in these endeavors has been limited and a consensus as to the best method is lacking.

Our approach to dealing with C was quite different. As cognitive scientists, we viewed the variability of C as the inevitable consequence of the sort of problem presented to the observers, i.e., create and maintain a consistent mapping from P to R. The problem of creating a mapping may or may not be difficult but it is definitely open ended, with very few constraints on the solution. Also, the problem of maintaining the mapping once it has been created could tax the limits of working memory. In fact, Petrov and Anderson (2000) and Petrov (2001) were able to model a number of different bias effects associated with various factors using the ACT-R (Anderson & Lebiere, 1988) architecture to model the memory processes involved. Based on this view, our approach has been to attempt to eliminate these effects by providing computerized support for establishing and maintaining the scale.

Constrained Scaling

Constrained scaling is a form of magnitude estimation (i.e., observers report numbers to match stimulus values). The goal of constrained scaling is to calibrate observers to the same C function before scaling the stimulus dimension of interest, similar to the way that physical measuring instruments are calibrated before use (Ward, 1991). Constrained scaling (West, Ward, & Khosla, 2000) is based on four claims about C: (1) that C is cognitively penetrable, (2) that C is heavily influenced by ad hoc decisions made early in the scaling process, (3) that the C process makes heavy demands on working memory which leads to instability across the task, and (4) that C is independent of the perceptual modality being judged (i.e., if the perceptual modality is changed it does not directly cause a change in C, although an interruption in the process could disrupt and indirectly alter C). Provided these assumptions are true, it should be possible to train observers to use a predetermined C function, and to support the maintenance of it in memory by refreshing it through a computerized feedback system.

Constrained scaling involves two phases, a learning phase and a test phase. In the learning phase, feedback is used to train observers to respond to a standardized set of stimulus magnitudes according to a predetermined response scale. This is done across several trials by presenting learning set stimuli and having the observer rate the perceived magnitude by entering an R value. On the interface we have been using this can be done by entering a value in a text box or by using a specially designed scroll bar that allows the observer to move the slider by units of 10, 1, 0.1, and 0.01. The scroll bar runs from 0 to 100 (although the observers are instructed that they may enter R values above 100). After this the observer clicks a button marked, "OK," and their R value is replaced with the correct R value. The point of this is to build C functions that are the same across observers and to give them the practice they need to become familiar with it. Provided that P is highly similar across observers, training the observers so that they all correspond to the same function relating S and R, implies they have the same C function, although it is possible that the details of how they cognitively implement and maintain the C function may differ.

The choice of the scale to be learned should be based on learnability and the mathematical desirability of the scale. Similar to West et al (2000), we used a power function with an exponent similar to what would be found using ME (i.e., we accept, to some extent, Stevens' argument that free ME produces scales that people find more natural to use) and K was set so that the scale range was approximately from 1 to 100 (as we believe this is a range that people are familiar with).

Research has shown that, with feedback on each trial, people can learn these scales quite accurately (King & Lockhead, 1983; Koh & Meyer, 1991; Koh, 1993; West & Ward, (1994); Marks, Galanter, & Baird, 1995). However, we have found that once the feedback is taken away, people start to drift off of the learned scale. Therefore, during the test phase the learned scale is presented on every second trial followed by feedback, so that the form of the scale is constantly refreshed in memory. On the alternate trials, test stimuli, different from the learned stimuli, are presented without feedback. The observers are instructed to use the learned scale to respond to the test stimuli as well as the learned stimuli. They are also told that the response range of the test stimuli may be greater or less than the response range of the test stimuli.

This general approach was used in West et al (2000) and the results were compared to other psychophysical methods. In that study, the learned scale stimuli were 1000 Hz tones between 32 dB and 99 dB, spaced at 1 dB intervals. The learned scale responses were numbers from 1 to 100 related to the stimulus magnitudes by a power function with an exponent of 0.600 (taken from the International Organization for Standardization, 1959). The test stimuli were 65 Hz tones and light brightness. The results, a full discussion of the psychophysical meaning of the results, and a comparison to other methods is presented in West et al (2000). Here we will just point out that constrained scaling produced very low levels of inter-observer variability compared to ME and CMM. Furthermore, the only method that we could find that produced similar low levels of inter-observer variability was conjoint measurement as applied to combined pairs of tones (Schneider, 1988). However, this methodology exploits the fact that, under the right conditions, loudness is additive for two tone combinations, which limits its application to auditory stimuli. It also requires a large number of trials.

Scaling Video Frame Rates

The results from West et al (2000) clearly demonstrated that training observers and using external means to constantly refresh their memory produces highly reliable scaling results. This indicates that arbitrary decisions about how to structure a scale and insufficient resources for maintaining the scale in memory are the primary source of inter-observer variability in direct scaling. However, it was still unclear how observers use the feedback to maintain a representation of the scale. We speculated that observers memorized a limited number of perceived magnitude/response pairs and interpolate to get responses in-between (see Ward & West, 1988, for an example of people using this strategy in a similar type of task). If this is the case then constrained scaling should work if the observers are only supplied with feedback on a limited number of S/R pairs instead of many pairs covering the whole range (as in West et al 2000).

We applied this methodology in a study designed to look at the effect of content type on the perception of frame rate in video clips. Specifically, we were interested in whether or not speed of movement in the clip alters the perception of frame rate. To do this we began with a pilot study using magnitude matching. Magnitude matching is a version of ME in which two different stimuli are alternately presented in the same scaling task (J. C. Stevens & Marks, 1980). In this case we used a fast paced video clip and a slow paced video clip. The results, averaged across observers, indicated that the exponent for frame rate was approximately 0.90. No

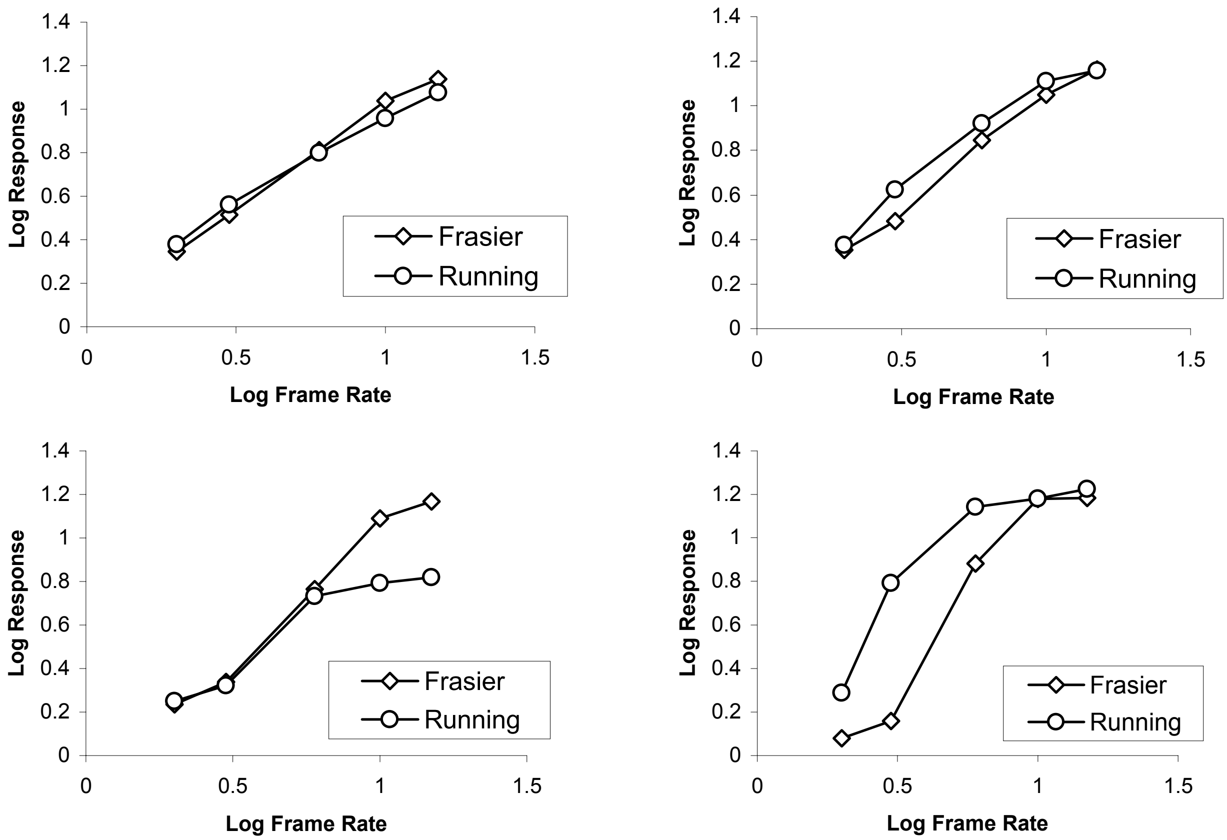

Figure 1. Psychophysical functions for four representative observers. The top row shows two observers who obeyed the power law and the bottom row shows two observers who deviated from it

significant effects for content were found (Boring, West, & Dillon, 2000).

For the constrained scaling experiment we used only five stimulus levels for training (2, 3, 6, 10, and 15 frames per second). Observers were taught, using feedback, to respond to these frame rate levels according to a power function with an exponent of 0.90. The observers were given 50 trials to learn the scale and the stimuli were presented randomly. The content of the video clip was moderate in speed (medium speed hip hop dancing).

During the test phase, the observers were instructed that the same hip-hop clips would be presented with feedback on every second trial, and that on the alternative trials a different video clip would be presented. The Observers were told to respond to the other clip using the same scale they learned for the hip-hop clip, but that the frame rate levels would not necessarily be the same and that there would be more than five versions of the new clip. This was actually not true; the test stimuli were generated using the same frame rate levels as the learning stimuli. However, the observers did not know this as the stimuli were spaced less than one JND (just noticeable difference) apart. We mislead our observers so that they would be open to responding with the whole range of responses. The observers all completed two test phase sessions, one using a fast content clip (children running) and one using a slow content clip (a clip from the Fraser show of Fraser talking). The order of the sessions was counterbalanced and another 50 trials of training were presented in-between. All stimuli were presented in random order.

Results

As in West et al (2000), we found that constrained scaling did not produce outliers, so we used mean response values for scaling the responses instead of medians. From a visual inspection of the graphed functions from the test phase trials it was clear that four observers produced functions with relatively large nonlinear trends (see Figure 1). This is actually not uncommon in ME (Luce, & Mo, 1965). The normal procedure would be to throw them out or to average across them, along with the functions of the other observers. However, since we are interested in individual differences, we note that these four were less able than the other observers to exploit the external scaling aids offered by constrained scaling. This indicates that individual differences in strategy, cognitive ability, and/or effort still play a role. Since these deviations were not unusually large by ME standards we analyzed the data both with them in and with them out. The remaining six observers produced functions that could reasonably be treated as linear (see Figure 1).

West et al reviewed 14 studies that provided individual observer results for ME and CMM, and calculated the standard deviation divided by the mean for the individual exponent values from each study. As a basis for comparison we took these values and calculated the mean, which was 0.333, the standard deviation, which was 0.080, and the 0.05 confidence interval, which was plus or minus 0.042. Even with the four linearly deviant observers included, the mean of the individual exponent values divided by the standard deviation was 0.190 for the Fraser clip and 0.150 for the children running clip, significantly lower than what would be expected with ME or CMM. Without the four deviants included, the mean divided by the standard deviation was 0.076 for the Fraser clip and 0.047 for the children running clip. These values were similar to the mean divided by standard deviation values found by West et al (2000) using constrained scaling (these values were 0.045, 0.066, and 0.152).

Also, because of the low variability we were able to detect a small but significant difference in exponent values both with ($P < 0.01$) and without ($P = 0.01$) the four linearly deviant observers, indicating that the exponent values for the slower video were higher than the exponent values for the faster video. This finding illustrates the advantage of having more precise ways of measuring perceived magnitudes (note, since the purpose of this paper is to examine the cognitive aspects of scaling, we will not discuss why this difference might exist).

Discussion

These findings replicate the West et al (2000) finding that augmenting the cognitive abilities of the observer can significantly reduce inter-observer variability and, more generally, supports the four theoretical assumptions behind constrained scaling (see above). The results also support the hypothesis that people can maintain scales in memory by memorizing a limited number of S/R pairs. By providing support to remember five S/R pairs we significantly reduced inter-observer variability to a level comparable to that found in West et al (2000), who provided feedback for a large number of responses. Other strategies may also be possible but, at the very least, this result shows that providing support for remembering a small number of S/R pairs can provide a significant advantage.

In terms of strategy, examining the actual responses that the observers made revealed that they took a category scaling approach. Two observers used the five R values they had learned almost exclusively. The other observers added only a few new R values and some stopped using one or two of the learned R values. The new R values also tended to be used as categories, that is, they were used repeatedly. This was quite different from the West et al (2000) observers who responded with a wide range of R values. From this it would appear that observers prefer to continue using a response strategy that resembles the one they were trained on. This may be due to observers inferring that the number of test stimuli will be similar to the number of learning stimuli, or it may be that teaching them to respond in a particular way creates cognitive structures that are not amenable for doing the task in other ways.

The fact that observers were able to respond accurately using a category scaling strategy, on a scale that was determined using ME, suggests that training and providing feedback to observers eliminates the factors that cause category scaling to produce different results from ME. This result is quite promising as it suggests that providing external support for the scaling process can wipe out methodologically induced biases.

Conclusions

These results provide compelling evidence that cognitively augmenting observers can substantially increase the reliability of psychophysical scaling, which is particularly important for measuring and studying individual differences and small group differences (as in this study). We also believe that this approach will eventually provide a means for assessing the validity of the scales as well. This is based the assumption that the further a learned scale is from the natural scale, the more cognitive resources will be required to maintain the mapping (C) from P to R (for some evidence of this see Marks, Galanter, & Baird, 1995; West et al, 2000). To improve further we need to better understand the strategies available to observers, and how to more effectively intervene to support the scaling process. Eventually, we hope that this approach will lead to psychophysical measurement techniques that have the same unambiguous status as physical measuring techniques.

References

Algom, D., & Marks, L. E. (1984). Individual differences in loudness processing and loudness scales. *Journal of Experimental Psychology: General*, *113*, 571-593.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berglund, M.B. (1991). Quality assurance in environmental psychophysics. In S.J. Bolanowski Jr.& G.A. Gescheider (Eds.) *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens.* Hillsdale, New Jersy: Lawrence Erlbaum Associates, Publishers.

Bolanowski, S. J., & Gescheider, G. A. (1991). *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Boring, R.L., West, R.L., & Dillon, R.F. (2000). Evaluation of framerate quality for different video content types. Poster presented at the CITO Digital Media Research Review, Toronto, Ontario, February 15, 2000.

International Organization for Standardization (1959). *Expression of physical and subjective magnitudes of* *sound [ISO/R-131-1959(E)]*. Geneva: International Organization for Standardization.

King, M. C., & Lockhead, G. R. (1981). Response scales and sequential effects in judgement. *Perception & Psychophysics*, *30(6)*, 599-603.

Koh, K. (1993). Induction of combination rules in two dimensional function learning. *Memory and Cognition*, *21(5)*, 573-590.

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *17(5)*, 811-836.

Logue, A. W. (1976). Individual differences in magnitude estimation of loudness. *Perception & Psychophysics*, *19(3)*, 279-280.

Luce, D. R., & Mo, S. S. (1965). Magnitude estimation of heaviness and loudness by individual observers: A test of a probabilistic response theory. *The British Journal of Mathematical and Statistical Psychology*, *18(2)*, 159-174.

Marks, L. E. (1974). On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. *Perception & Psychophysics*, *16(2)*, 358-376.

Marks L. E. (1991). Reliability of magnitude matching, *Perception & Psychophysics*, *49(1)*, 31-37.

Marks, L. E., Galanter, E., & Baird, J. C. (1995). Binaural summation after learning psychophysical functions for loudness. *Perception & Psychophysics, 57*, 1209-1216.

Marks, L. E., & Stevens, J. C. (1965). Individual brightness functions. *Perception & Psychophysics*, *1*, 17-24.

Petrov, A. (2001). Fitting the ANCHOR model to individual data: A case study in Bayesian methodology. Fourth international conference on Cognitive modeling (pp. 175-180). Hillsdale, NJ: Lawrence Erlbaum Associates.

Petrov, A. & Anderson, J. R. (2000) ANCHOR: A memory based model of category rating. *Proceedings of the 22nd annual confernce of the cognitive science society* (pp. 369-374). Hillsdale, NJ: Lawrence Erlbaum Associates.

Poulton, E. C. (1989). *Bias in quantifying judgements*. London: Lawrence Erlbaum Associates, Publishers.

Rule, S. J., & Markely, R. P. (1971). Subject differences in cross-modality matching. *Perception & Psychophysics*, *9*, 115-117.

Schneider, B. (1980). Individual loudness functions determined from direct comparisons of loudness intervals. *Perception & Psychophysics*, *28*, 493-503.

Schneider, B. (1988). The additivity of loudness across critical bands: A conjoint measurement approach. *Perception & Psychophysics, 43,* 211-222.

Stevens, J. C. (& Marks, L. E. (1980). Cross-modality matching functions generated by magnitude estimation. *Perception & Psychophysics*, 27, 379-389.

Stevens, S. S. (1956). The direct measurement of sensory magnitudes – loudness. *American Journal of Psychology, 69*, 1-25.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural and social Prospects*. New York: A Wiley-Interscience Publication.

Stevens, S. S. (1971) Issues in psychophysical measurement. *Psychological Review, 78, 5*, 426-450.

Teghtsoonian, M., & Teghtsoonian, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics, 33*, 203-214.

Ward, L. M. (1991). Associative measurement of psychological magnitude. In S. J. Bolanowski & G. A. Gescheider (Eds.), *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens* (pp. 79-100). Hillsdale, New Jersy: Lawrence Erlbaum Associates, Publishers.

Ward, L. M., & West, R. L. (1998). Modelling human chaotic behaviour: Non-linear forecasting analysis of logistic iteration. Nonlinear Dynamics, Psychology, and Life Sciences, 2, 4, 261-281.

West, R. L., & Ward, L. M. (1994). Constrained Scaling. In L. M. Ward (Ed.) *Fechner Day 94*. Vancouver: International Society for Psychophysics.

West, R. L., & Ward, L. M. (1998). The value of money: Constrained scaling and individual differences. Fechner Day: Proceedings of the Fourteenth Annual Meeting of the International Society for Psychophysics.

West, R. L., Ward, L. M., & Khosla, R. (2000). Constrained scaling: The effect of learned psychophysical scales on idiosyncratic response bias. Perception & Psychophysics, 62(1), 137-151.

Wanschura R. G., & Dawson, W. E. (1974). Regression effect and individual power functions over sessions. *Journal of Experimental Psychology, 102(5)*, 806-812.

West, R. L., & Ward, L. M. (1994). Constrained Scaling. In L. M. Ward (Ed.) *Fechner Day 94*. Vancouver: International Society for Psychophysics.

Zwislocki, J. J., & Goodman, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics, 28*, 28-38.