THE UNIVERSITY OF

WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

This paper is made available online in accordance with
publisher policies. Please scroll down to view the document
itself. Please refer to the repository record for this item and our
policy information available from the repository home page for
further information.

To see the final version of this paper please visit the publisher's website.
Access to the published version may require a subscription.

# Relative Judgment and Knowledge of the Category Structure

Neil Stewart and William J. Matthews
University of Warwick

For evenly spaced stimuli, a purely relative judgment account of unidimensional categorization performance is trivial—all that is required is knowledge of the size of stimulus difference corresponding to the width of a category. For unevenly spaced stimuli, long-term knowledge of the category structure is required. We argue that such knowledge does not necessitate a direct, absolute mapping between (representations of) stimulus magnitudes and category labels. We show that Stewart, Brown, and Chater's (2005) relative judgment model can account for data from absolute identification experiments with uneven stimulus spacing.

Models of the identification and categorization of simple perceptual stimuli have been divided into two classes. First, there are those models in which there is a direct mapping between stimulus magnitudes or regions of stimulus space and category labels (Ashby & Townsend, 1986; Durlach & Braida, 1969; Kent & Lamberts, 2005; Lacouture & Marley, 2004; Luce, Green, & Weber, 1976; Marley & Cook, 1986; Nosofsky, 1986, 1997; Petrov & Anderson, 2005). Second, there are relative judgments models in which there is no long-term mapping between stimulus magnitudes or regions of stimulus space and category labels. These mappings are assumed to be unavailable or, at least, unused. Instead, in relative judgment models, judgments are made with reference to recently encountered stimuli and category labels (Holland & Lockhead, 1968; Laming, 1984, 1997; Stewart, 2007; Stewart & Brown, 2004; Stewart, Brown, & Chater, 2002; Stewart et al., 2005). Most recently, the SAMBA model integrates these two accounts (Brown, Marley, Donkin, & Heathcote, 2008).

Recently Brown, Marley, Dodds, and Heathcote (in press) discussed how relative judgment might account for data from category structures where stimuli are unevenly spaced. Figure 1 illustrates the three spacings of stimuli along a single continuum taken from their experiment. The experiment was absolute identification, in which stimuli are presented one at a time (with replacement) and participants are required to identify each stimulus with a unique label. Trial-by-trial feedback of the correct answer is provided. A
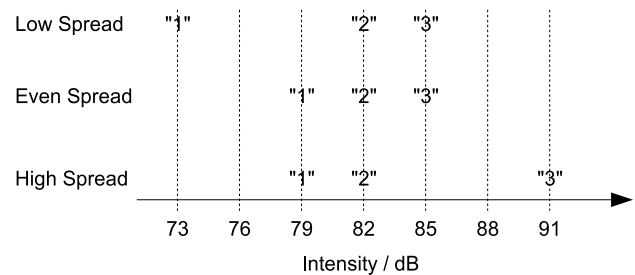
*Figure 1.* The three category structures used by Brown et al. (in press). Numbers in quotation marks are category labels.

relative-judgment account for the Even-Spread Condition is quite straightforward. All that is required is knowledge of the exchange rate between stimulus units and response units. Thus, if one has learned that 3 dB equals one response category then, given the category of the previous stimulus and a perception of the difference between current and previous stimuli, one can determine the category of the current stimulus (see Holland & Lockhead, 1968, for a relative judgment account of this sort).

When stimuli are not evenly spaced along the continuum, as in the Low- and High-Spread Conditions, relative judgment is not quite so straightforward. Now, as Brown et al. (in press) describe, the above strategy will not work. There is no simple exchange rate between stimulus differences and response scale differences because the size of stimulus difference required for a change in response category varies as a function of the previous stimulus.

In the remainder of this article, we describe a look-up table model of relative judgment. We show that, if redundancy in the category structure representation is exploited, the look-up table representation is equivalent to the Stewart et al. (2005) relative judgment model (known by the acronym RJM). Finally, we demonstrate that this model can account for the uneven spacing data from Brown et al. (in press).
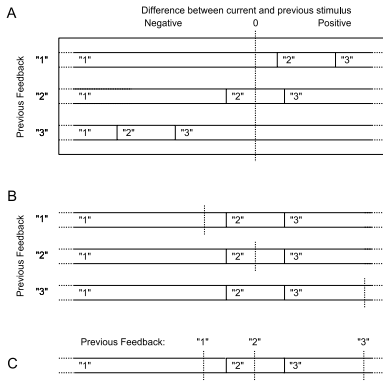
*Figure 2.* A: A look-up table relative judgment model. B: The same look-up table aligned by category boundaries rather than by zero difference. C: The same information represented on a single response scale. Digits in quotes on the left hand side of each region refer to the linguistic category label for that region. Dashed lines denote the zero points and correspond to $z_{\text{“1”}}$, $z_{\text{“2”}}$, and $z_{\text{“3”}}$. The horizontal axis represents the difference between consecutive stimuli and corresponds to the $\frac{D^C_{n,n-1}}{\lambda}$ term in Equation 3.

## A Look-Up Table Model of Relative Judgment

Consider a purely relative judgment model, in which one constructs a look-up table for the current response as a function of (a) the previous feedback linguistic label and (b) the difference between the previous stimulus and the current stimulus. Figure 2A shows an example of such a look-up table. For example, when the previous feedback is "2", then small differences will lead to responding "2", larger negative differences will lead to responding "1", and larger positive differences will lead to responding "3". This look-up table might be learned over the initial trials of the experiment. If stimulus differences are perceived and remembered accurately, then performance could be perfectly accurate.

The look-up table model could be considered as an exemplar model in which the exemplars are {previous feedback label, stimulus difference} pairs and the exemplars are stored with their category labels. Rather than pursue this possibility, we show below that Stewart et al.'s (2005) RJM can be interpreted within this look-up table framework.

## Exploiting Redundancy in the Look-Up Table

Because stimuli are drawn from a single continuum with a static category structure (i.e., there is a consistent mapping between stimuli and feedback in the experiment), there is redundancy in the look-up table representation. Across the different levels of the previous feedback label, a given category must have the same width. For example, in Figure 2A, the width of Category "2" must be the same, irrespective of the previous feedback. Figure 2B shows the same table aligned by the start and end of the Category "2" regions. The dashed lines represent the zero points, but they are now no longer aligned with one another. These *zero points* are essentially a record, for different possible previous feedback labels, of where to begin the look up of stimulus differences.

In general, the look-up table represents the location of category boundaries on a stimulus difference scale relative to the location of a zero difference (Figure 2A). Alternatively, one can represent the location of the zero points relative to the category boundaries (Figure 2B). In both interpretations the same information—the relative locations of criteria and zero points—is represented. Figure 2C shows these three points represented on a single scale. Note how each zero point is associated with a linguistic feedback label. Again, the information represented is the same as in Figures 2A and 2B. This scale is the internal response scale from the Stewart et al. (2005) RJM, as we describe below.

## Equivalence to the Stewart et al. (2005) RJM

Here we briefly summarize the RJM to make the look-up table interpretation clear (but see Stewart et al., 2005, for details). In the RJM, perceptual representations of the absolute magnitudes of stimuli are unavailable or unused. Instead, the elemental perceptual unit is the difference between the current stimulus and the previous stimulus, $D_{n,n-1}$ (Equation 1). (Of course, there must be some peripheral representation of absolute stimulus magnitude to allow the differences between consecutive stimuli to be apprehended over silent or blank intervals between stimuli, but it is only stimulus differences that are utilized in identification.)

$$D_{n,n-1} = A \ln\left(\frac{X_n}{X_{n-1}}\right) \qquad (1)$$

The units of $D_{n,n-1}$ are arbitrary, and depend on the sensory constant $A$. $X_n/X_{n-1}$ is the ratio of the physical magnitudes of the current and previous stimuli. Consecutive differences are assumed to be confused with one another, producing a contaminated estimate of the difference between the current and previous stimulus.

$$D^C_{n,n-1} = \sum_{i=0}^{n-2} \alpha_i D_{n-i,n-i-1} \qquad (2)$$

where $\alpha_0 = 1$ and $\alpha_i > \alpha_{i+1}$ for all $i > 0$.

The stimulus difference is used to derive a distribution for the estimate of the magnitude of the current stimulus on an internal response scale, $\mathbf{R}_n$.

$$\mathbf{R}_n = z_{n-1} + \frac{D^C_{n,n-1}}{\lambda} + \rho\mathbf{L} \qquad (3)$$

$z_{n-1}$ is the location on the internal response scale of the zero point associated with $F_{n-1}$ (the linguistic category label from the previous trial). The $\rho\mathbf{L}$ term represents noise. $\mathbf{L}$ is

the Laplace distribution, with mean 0 and scale parameter $\sigma$ (Stewart, 2007). $\rho$ is the range of available responses given the previous feedback and the difference between the current or previous stimuli (see Stewart et al., 2005, for details). $\lambda$ is the size of stimulus difference corresponding to an (arbitrary) unit on the internal response scale.

Note here that we have changed the notation from the 2005 version to make clear the distinction between the category label and the associated location on the internal response scale. To see why the distinction is crucial, consider an absolute identification experiment in which stimulus labels are not integers but are instead nonsense words (e.g., "deg" for Stimulus 1, "rop" for Stimulus 2, ...). The participant must learn the locations on the internal response scale associated with each nonsense word. In the 2005 version of Equation 3, $F_{n-1}$ appeared in place of $z_{n-1}$. Because the units of the internal response scale are arbitrary, using the integers to represent the locations associated with the feedback labels "1", "2", ... on the internal response scale is acceptable for evenly-spaced stimuli. For nonsense words this makes no sense: One cannot add a stimulus difference to "rop"; "rop" must be converted into some magnitude on the internal response scale. The 2005 simplification of using integers for the internal magnitudes is further confusing because it suggests that the internal response scale actually is the set of integers used as feedback labels, when in fact the units of the internal response scale are arbitrary. And the 2005 simplification is inadequate because it does not distinguish between the verbal labels paired with stimuli by the experimenter and the representations of the points on the internal response scale associated with those labels by the participant. In making the substitution of $z_{n-1}$ for $F_{n-1}$ here, the workings of the model are clarified but the predictions of the model are unchanged.

A core assumption of the model is that a stimulus cannot give rise to activation on the internal response scale except via the relative judgment process in Equation 3. That is, there is no mapping between raw representations of stimulus magnitudes and locations on the internal response scale (cf. Brown et al.'s, 2008, assumption that their rehearsal scale is not addressable with category label tags).

The probability of responding with each category label is given by dividing the internal response scale with optimally located criteria, $x_{\text{"1"}-\text{"2"}}, x_{\text{"2"}-\text{"3"}}, \ldots, x_{\text{"}N-1\text{"}-\text{"}N\text{"}}$ where there are $N$ categories. Subscripts refer to the category labels of the regions a criterion divides. In the 2005 treatment of evenly-spaced stimuli, Stewart et al. used a James-Stein estimator to approximate the optimal locations. For unevenly-spaced stimuli, criteria are assumed to be placed to maximize accuracy. One interpretation of the division of the internal response scale has $\mathbf{R}_n$ representing a distributed pattern of activation across a population of neurons, with the $x_{\text{"}i\text{"}-\text{"}i+1\text{"}}$ criteria representing the limits of the receptive fields of neurons summing activation in the population.

Figure 2C represents the internal response scale in the RJM. The dashed lines represent zero points $z_{\text{"1"}}$, $z_{\text{"2"}}$, and $z_{\text{"3"}}$ associated with each category label. For example, if $F_{n-1} = $ "3" then $z_{n-1} = z_{\text{"3"}}$. The category boundaries repre-sent the criteria $x_{\text{"1"}-\text{"2"}}$ and $x_{\text{"2"}-\text{"3"}}$. The RJM representation of the category structure contains the same information as the look-up table model.

We think it is uncontroversial to claim that judgment based on the look-up table is relative and not absolute. Partitioning the table up into labeled response regions does not imply some direct link between response category labels and raw stimulus magnitudes or representations of stimulus magnitudes. Thus we also claim that judgment in the RJM is relative and not absolute. Having a representation of the category structure does not mean that there must be a mapping between raw stimulus magnitudes or representations of stimulus magnitudes and the labels assigned to each stimulus. In the RJM there is a direct mapping or transformation between the internal response scale and category labels (i.e., the category structure), but there is no direct mapping between the internal response scale and stimulus magnitudes. Stimulus magnitudes only cause activation on the internal response scale via the relative judgment process.

More generally, even when judgment is entirely relative, it is still possible to learn something about the category structure. For example, consider the Low-Spread Condition in Figure 1. Suppose you hear a stimulus and are told it is "1". Then suppose you hear the next stimulus which sounds much louder and are told it is "2". Then suppose you hear a third stimulus which sounds only a little bit louder and are told it is "3". By now you know that "2" is much louder than "1" and that "3" is only a little bit louder than "2". You've already got quite a good idea (better than just ordinal) about the stimulus structure derived from only relative judgment. Learning this structure doesn't mean you have somehow created a direct mapping between stimulus magnitudes and category labels. With accurate relative judgment there could even be an accurate (but scale-free) representation of the category structure.

## RJM Fits to the Brown et al. (in press) Experiment.

In fitting the RJM to Brown et al.'s (in press) experiment, the model had three free parameters: $\alpha_1$ representing confusion between the current and previous difference ($\alpha_i = 0$ for all $i > 2$), $\sigma$ representing noise in the mapping process, and $\lambda$ representing the size of stimulus difference equal to a single (and arbitrary) internal response scale unit. The location of criteria on the internal response scale $x_{\text{"1"}-\text{"2"}}$ and $x_{\text{"2"}-\text{"3"}}$ were arbitrarily fixed at 0 and 1 (without loss of generality, one may define the locations of the lower and upper bounds for response "2" as 0 and 1 respectively). The locations of the zero points $z_{\text{"1"}}$, $z_{\text{"2"}}$, and $z_{\text{"3"}}$ are not free parameters, but are assumed to be optimally located and were adjusted to maximize accuracy. That is, for every step in the best fitting of $\alpha_1$, $\sigma$, and $\lambda$ to the data, $z_{\text{"1"}}$, $z_{\text{"2"}}$, and $z_{\text{"3"}}$ were adjusted to maximize accuracy (not fit to the data) in a nested maximization. The zero points were different for each experimental condition because the zero points represent the category structure and the structure varied between experimental conditions.

Figure 3 shows the best-fit of the RJM. Best fitting parameters were $\alpha_1 = 0.197$, $\sigma = 0.137$, and $\lambda = 4.224$. The RMSE

is 0.066 and the greatest mismatch is 0.15, indicating a similar fit to that reported by Brown et al. (in press) for SAMBA (0.059 and 0.12, respectively). Mispredictions occur because the RJM is predicting accuracy that is too high for the lowest stimulus in the Low-Spread Condition and the lowest stimulus in the High-Spread Condition. However, for the corresponding stimuli in the mirror image conditions, the fit is quite accurate (i.e., accuracy is higher for the highest stimulus in the High-Spread Condition and the highest stimulus in the Low-Spread Condition). (Parker, Murphy, & Schneider, 2002, present some evidence for asymmetry between High- and Low-Spread Conditions in auditory intensity judgments, though their gain control hypothesis does not explain the full pattern here and would be harder to extend to nonprothetic dimensions.)

The RJM achieves this fit with three free parameters, though we would have considered relaxing the constraint $\chi = 1.0$ if necessary (see Stewart et al., 2005, for details). In fitting SAMBA four parameters were varied, with other values taken from fits to previous data sets (and presumably these parameters would have been varied too if necessary). Figure 4 shows the optimal locations of the zero points relative to the criteria in the RJM for the three experimental conditions.

## Judgment is Still Relative

Brown et al. (in press) argue that, generally, relative judgment models cannot predict data from uneven-stimulus-spacing designs. We offer the look-up table model as a counterexample of a purely relative judgment model which does account for data from uneven stimulus-spacing-designs.

Brown et al. (in press) argue that, specifically, the RJM cannot account for data from uneven-stimulus-spacings designs. Brown et al. (in press) present an extended RJM which can fit the data. In this model there is a mapping between feedback labels and (peripheral representations of) stimulus magnitudes. Like Brown et al. (in press), we would not describe the model as "relative" because there is a consistent, long-term mapping between stimulus magnitudes and feedback labels.

In our implementation there is no mapping between feedback labels and stimulus magnitudes. Instead, there is a mapping between feedback labels and the zero points on the internal response scale. This difference is crucial. Here, as in Stewart et al. (2005), there is long-term representation of the category structure. The locations of the zero points (and of the criteria) represent the category structure and are thus correlated with stimulus magnitudes. But the zero points are ideal-observer values acquired from only stimulus difference information. Zero points are not stimulus magnitudes. Information about the mapping between stimulus magnitudes and points on the internal response scale is assumed to be unavailable or unused. That is, presentation of a stimulus does not result directly in activation on the internal response scale. Activation is via the relative judgment process only.

To illustrate this point, consider Experiment 2 from Stewart et al. (2005), in which participants were asked to identify
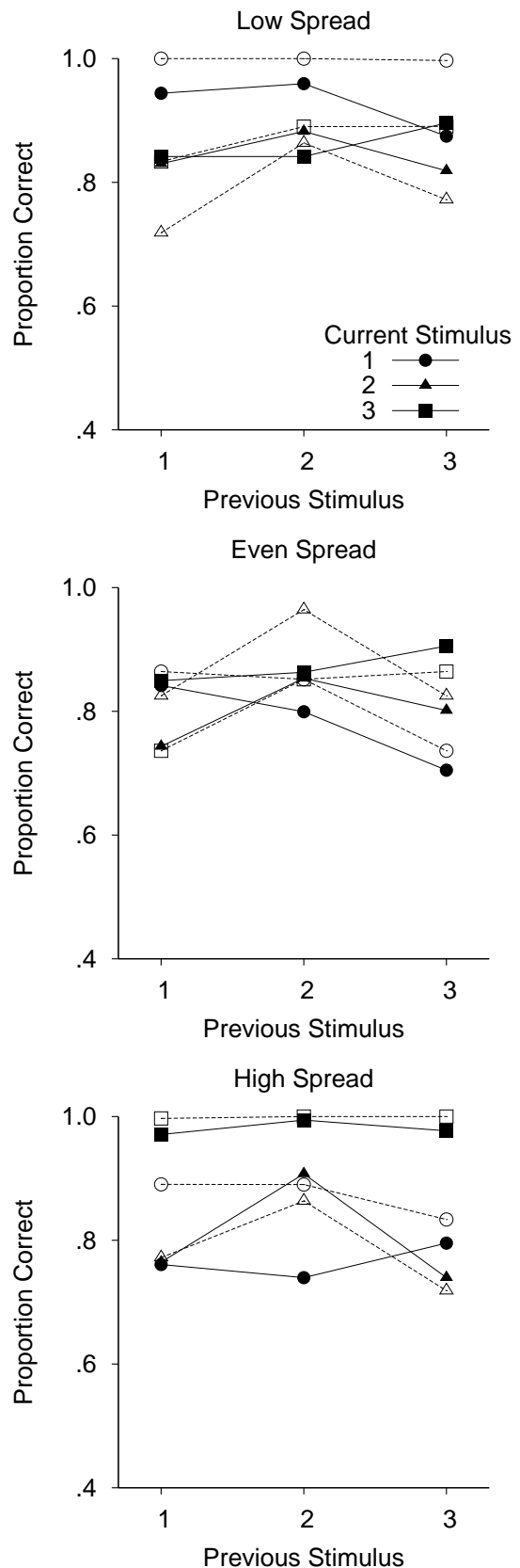


*Figure 3.* Fits of the RJM to the data from Brown et al.'s (in press) experiment. Solid symbols show data, open symbols show model predictions.
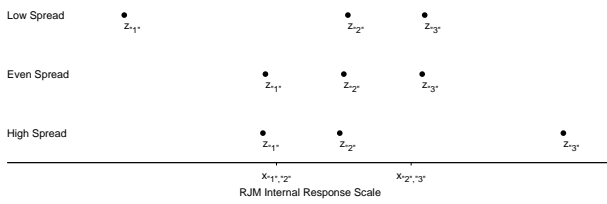
*Figure 4.* Optimal locations of criteria and zero points. $x_{\text{"1"–"2"}}$ and $x_{\text{"2"–"3"}}$ were arbitrarily fixed at 0 and 1. Note locations of zero points are set to maximize accuracy (reflecting their optimal nature) and not fit to the data.

10 evenly spaced tones varying in frequency. On a small number of trials participants were given misleading feedback. For example, the third highest stimulus might be given the feedback label "4". On the following trial, participants were found to be systematically biased to respond in the direction of the lie. Continuing the example, if the next stimulus was the fourth highest, participants had an increased tendency to overestimate the stimulus and respond "5" instead of the correct answer "4". This is exactly what the RJM predicts, because the misleading feedback will invoke the incorrect zero point (in the example, $z_{\text{"4"}}$ instead of $z_{\text{"3"}}$). In terms of the look-up table account, the misleading feedback will cause participants to map the stimulus difference onto response categories using the wrong previous-feedback row (in the example, the row for "4" instead of the row for "3"). If judgment were absolute and the response scale directly addressable from stimulus magnitudes, this effect would not be predicted. (See Brown et al., 2008, for an explanation using SAMBA, in which the mapping between magnitude estimates and response accumulators is completely re-adjusted to accommodate the misleading feedback.)

## Conclusion

We have shown that a relative model, the RJM, can account for data from experiments with uneven stimulus spacing. The RJM is a relative model in the sense that the fundamental unit submitted for judgment is the difference between the current stimulus and the last, rather than the position of the stimulus in the experimental range (e.g., Marley & Cook, 1984) or on a continuum that is independent of the other stimuli in the ensemble (e.g., Kent & Lamberts, 2005; Treisman, 1985). Just because a model is relative in this sense does not mean that the participant is without any kind of long-term knowledge of the category structure: Such knowledge can be obtained from stimulus differences and verbal feedback, without there being any direct association between the sensation evoked by an individual, isolated stimulus and the corresponding response label.

## References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Brown, S. D., Marley, A., Dodds, P., & Heathcote, A. (in press). Relative models do not provide a general account of absolute identification. *Psychonomic Bulletin & Review*.

Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. J. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115*, 396–425.

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*, 372–383.

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

Kent, C., & Lamberts, L. (2005). An exemplar account of the bow and set size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 289–305.

Lacouture, Y., & Marley, A. A. J. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & Psychophysics*, *66*, 1206–1266.

Laming, D. R. J. (1984). The relativity of "absolute" judgements. *British Journal of Mathematical and Statistical Psychology*, *37*, 152–183.

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

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

Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology*, *37*, 136–151.

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychological judgments applied to magnitude estimation. *Journal of Mathematical Psychology*, *30*, 339–390.

Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 347–365). Mahwah, NJ: Erlbaum.

Parker, S., Murphy, D. R., & Schneider, B. A. (2002). Top-down gain control in the auditory system: Evidence from identification and discrimination experiments. *Perception & Psychophysics*, *64*, 598–614.

Petrov, A. A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, *112*, 383–416.

Stewart, N. (2007). Absolute identification is relative: A reply to Brown, Marley, and Lacouture (2007). *Psychological Review*, *114*, 533–538.

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in categorizing tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 416–430.

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 3–11.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*, 881–911.

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties for a model of absolute judgment. *Journal of Mathematical Psychology*, *29*, 175–230.