Psychological Review
2007, Vol. 114, No. 2, 528–532

0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.2.528

Is Absolute Identification Always Relative? Comment on Stewart, Brown, and Chater (2005)

Scott Brown
University of California, Irvine

A. A. J. Marley
University of Victoria

Yves Lacouture
Université Laval

N. Stewart, G. D. A. Brown, and N. Chater's (2005) relative judgment model includes three core assumptions that enable it to predict accurately the vast majority of "classical" phenomena in absolute identification choices, but not the time taken to make them, including sequential effects, such as assimilation and contrast. These core assumptions, coupled with the parameter values used in the above-mentioned article, lead to the prediction that identification accuracy is low when a large stimulus on 1 trial is followed by a small stimulus on the next trial and vice versa. Data do not support this prediction. The authors identify a set of parameters that allow the model to better fit the data, but problems remain when the data are analyzed with a version of the discrimination measure (d') from signal detection theory. The fundamental problem is that the model fits data on average but at the expense of making incorrect predictions in detail.

Keywords: absolute identification, relative judgment theory, sequential effects

In *unidimensional absolute identification*, an observer must identify the presented one out of a possible N, $N = 2$ unidimensional stimuli. As Miller (1956) pointed out, a remarkable finding is that, as N approaches and then exceeds seven, an observer will make an accelerating proportion of errors. This is surprising given that the stimuli are typically chosen such that an observer can make very accurate *comparative* judgments between of any pair of the stimuli when the stimuli are presented alone. As Shiffrin and Nosofsky (1994) stated in an article reassessing the significance of Miller's classic article, "absolute identification has captured the imagination . . . not only because the empirical results are so startling but also because [they] provide perplexing problems for classic psychophysical models" (p. 358). Researchers continue to develop models for the classic results, with the main advance over the past decade being the careful analysis of data and models for full response time distributions. Stewart, Brown, and Chater (2005) provided an excellent review[1] of extant data and models for the choices made in absolute identification but not for the time required to make those choices. They went on to develop their relative judgment model (RJM) for the choices made in absolute identification but not for the time required to make them. The RJM focuses particularly on sequential effects, with the explanation of such effects being at the core of the model. Also, as the following summary indicates, the model falls within the stream of the study of absolute identification that emphasizes the theme "absolute judgment is relative."

There are three core assumptions to the RJM. The first is that the response to the current stimulus is selected relative to the context determined by the feedback (if given) on the previous trial, along with comparisons of the current stimulus with previous stimuli and responses (with the stimulus measure rescaled to the response scale). Second, it is assumed that the set of "available" (possible) responses on each trial is constrained by the relative ordering of the stimuli on the current and the previous trial. Third, the variability of the responses to a particular stimulus is assumed to be an increasing function of the number of available responses when that stimulus is presented. With appropriate parameter values, the first core assumption leads the RJM to predict *assimilation* and *contrast* effects. That is, the RJM correctly predicts that first-order temporal effects (i.e., those due to the stimulus and response on the previous trial) lead to assimilation (a bias toward the response on the previous trial) and that second- and higher-order temporal effects

Scott Brown, Department of Cognitive Science, University of California, Irvine; A. A. J. Marley, Department of Psychology, University of Victoria, British Columbia, Canada; Yves Lacouture, Ecole de Psychologie, Université Laval, Quebec, Canada.

This research was supported by Australian Research Council Linkage International Project Grant LX0348125 to the University of Adelaide for Doug Vickers, A. A. J. Marley, Andrew Heathcote, Philip Smith, and Michael Lee; Natural Science and Engineering Research Council (NSERC) Discovery Grant 8124-98 to the University of Victoria for A. A. J. Marley; and NSERC Discovery Grant 0291-83 to Université Laval for Yves Lacouture.

Correspondence concerning this article should be addressed to Scott Brown, who is now at the School of Psychology, Aviation Building, University of Newcastle, Callaghan NSW 2308, Australia. E-mail: Scott.Brown@newcastle.edu.au

[1] Therefore, in the following, we are extremely selective in our citations and sincerely hope we offend no one by this selectivity.

lead to contrast (a bias away from stimuli and responses on those earlier trials).[2]

This comment focuses on the RJM's assumption that the variability of the representation used in reaching a decision on the current trial is a function of the number of available responses on that trial. To make the discussion concise, we introduce some notation. Let $S_k = i$ indicate that the stimulus with ordinal value i, $i = 1, \ldots, N$ was presented on trial k, and let $R_k = j$ indicate that response j, $j = 1, \ldots, N$ was made on trial k. For example, $S_n = 10$, $R_{n-1} = 8$ means that the stimulus with ordinal position 10 ("Stimulus 10") was presented on trial n and that Response 8 was made on trial $n - 1$.

Suppose that an observer is performing absolute identification with 10 stimuli with correct feedback provided. Also, suppose that the responses are associated with the stimuli in the natural order (i.e., the larger the stimulus, the larger the response). Now, suppose that $S_{n-1} = 3$ and $S_n = 5$. The RJM then assumes that the observer can accurately determine that Stimulus 5 is greater than Stimulus 3, so the only available responses on trial n are those greater than 3 (i.e., the possible values for $\mathbf{R}_n$ are 4, . . . , 10). Finally, the RJM assumes that, with these seven possible responses, the variability of the representation that determines the response is quite large, thus leading to relatively low accuracy. In contrast, suppose that $S_{n-1} = 3$, but it is followed by $S_n = 1$ on the current trial. Paralleling the above assumptions, the RJM assumes that the observer can accurately determine that Stimulus 1 is smaller than Stimulus 3, so the only possible values for $\mathbf{R}_n$ are 1 and 2. In this case, there are fewer possible responses, so the variability of the representation that determines the response is quite small, leading to relatively high accuracy.

Formally, these assumptions are represented in the RJM by Equations 3 through 5 in Stewart et al. (2005), which we briefly review here. Their Equation 4 is:

$$\mathbf{R}_n = F_{n-1} + (D^C_{n,n-1} / \lambda) + \rho\mathbf{Z}.$$

This equation states that distribution of the value of the response on trial $n(\mathbf{R}_n)$ is determined by three factors. The mean of this distribution is determined by the sum of the feedback (i.e., the numeral associated with the correct response) given on the prior trial (F_{n-1}) and a weighted average of the responses given on previous trials (the term $[D^C_{n,n-1} / \lambda]$, which is specified by Stewart et al.'s, 2005, Equation 3). Variability is introduced through the term $\rho\mathbf{Z}$, where $\mathbf{Z}$ is a normally distributed random variable with zero mean and unit standard deviation, and ρ is given by Stewart et al.'s Equation 5. That equation specifies that ρ increases linearly with the number of available responses, in the manner described above. On each trial, a response is simulated by generating a random sample from the distribution $\mathbf{R}_n$ and comparing its magnitude to $N - 1$ criteria that divide the response continuum into the N possible response categories. We implemented these assumptions in a simple computer program to simulate responses from the RJM (see below for details).

As we now show, the core assumptions of the RJM, with the parameter values presented by Stewart et al. (2005), lead to the prediction that accuracy will be low when a large stimulus on one trial is followed by a small stimulus on the next trial, or vice versa. However, we show that this prediction is contradicted by data from Lacouture (1997) that were considered by Stewart et al. (2005). Also, we discuss ways in which the model can be modified to avoid this problem—at least partially. We end with some discussion of implications for future analyses of absolute identification data and models.

Analyses

We first present a way of plotting probability correct data from absolute identification tasks that gives an indication of the nature of first-order temporal effects in this aspect of the data. Note that these plots show nothing about the distribution of errors across (incorrect) responses, even though the latter are important in determining the size of assimilation and contrast effects. Our first plot shows data averaged across trials in a way that masks some details, but later we break down the data in more detail.

For both data and theory, we calculated the average probability of a correct response conditional on the ordinal difference between the stimuli presented on the current and the previous trials (we adapted this analysis from Rouder, Morey, Cowan, & Pfaltz, 2004, Figure 5). The solid circles and dotted line in Figure 1 show this analysis carried out on the pooled[3] data consisting of the approximately 14,000 observations of the standard absolute identification experiment of Session 1 of Conditions 2, 3, 4, and 5 in Lacouture (1997). The error bars show ±2 standard errors, calculated assuming independent binomial distributions. This subset of Lacouture's data was analyzed by Stewart et al. (2005, Figure 5) for assimilation and contrast effects. They did not present, or fit, the data in the form of our Figure 1.

The x-axis in Figure 1 shows the ordinal difference between the current and prior stimuli. For example, the middle of the graph ($x = 0$) represents response accuracy when the current stimulus is identical to the prior stimulus. For our purposes, the most telling parts of the data plot in Figure 1 are the extreme ends. The left end of the graph shows response accuracy when the difference between the current and prior stimulus is −9 (i.e., when $S_n = 1$ and $S_{n-1} = 10$). Similarly, the right end of the graph shows response accuracy for a difference of +9 (i.e., when $S_n = 10$ and $S_{n-1} = 1$). The data show high accuracy (about 75%) for those pairs of stimuli. In fact, these are the largest accuracy values in this data plot, although the response accuracy for repeated stimuli (center of the graph) is also quite high (around 70%).

Next, we simulated 1,000,000 stimuli and responses from the RJM, according to details in Stewart et al. (2005). We used this large number of simulated responses so that the model predictions shown below are essentially errorless. We used the parameter values that Stewart et al. reported in their Table 3 as giving the best fits of the RJM to the assimilation and contrast effects in Lacouture's (1997) standard absolute identification task. The reader may

[2] Care is needed in determining the extent to which these effects depend on previous stimuli, as opposed to the previous responses and/or the correct responses for previous stimuli (Mori & Ward, 1995; Ward & Lockhead, 1971).

[3] These data are pooled across several participants. The problems associated with analyzing such pooled data can be distinguished from, although they are likely related to, the types of averaging—across trials—that are our focus in this note (see the Discussion section).

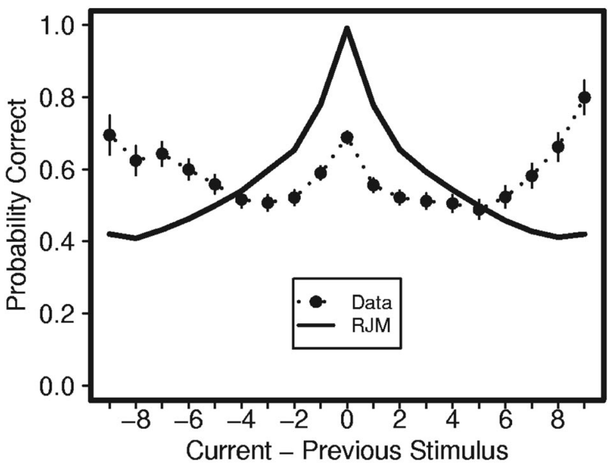

Figure 1. Average probability of a correct response as a function of the ordinal difference between the current and previous stimuli. Data are from Lacouture (1997), and model predictions are from the relative judgment model (RJM). The *x*-axis gives the difference between the current and previous stimuli. For example, a stimulus repetition gives a difference of $x = 0$; a change from one extreme stimulus to the other gives a difference of $x = +9$ or $x = -9$. Note that the RJM predicts poorest performance for the largest differences, but the data show greatest accuracy for those differences. We have omitted labels for odd-numbered differences on the *x*-axis (but not the corresponding data).

wish to check the accuracy of our simulation, so we have provided the code, with instructions and comments.[4]

The simulation results are shown by the solid line in Figure 1. It is clear that the RJM significantly overpredicts the average probability correct for small differences between the current and the previous stimulus and significantly underpredicts the average probability correct for larger differences. In particular, the RJM predicts poor performance (about 42% correct) when either extreme stimulus is followed by the other extreme stimulus (at the ends of the graph). This is a consequence of RJM's assumption that response variability increases with the relative number of available responses. In contrast, the RJM does make correct predictions for the center of the plot, at least qualitatively. The center of the plot represents repeated stimuli ($x = 0$ when a stimulus is directly repeated). In both the data and the model predictions, there is a peak in the center of the graph illustrating the accuracy bonus for repeated stimuli and smaller bonuses for near-repeats (such as when Stimulus 3 follows Stimulus 4).

Figure 1 shows that a prediction of the RJM with the given set of parameter values—low accuracy for one edge stimulus when it follows the other edge stimulus—is clearly not supported by the data. The reader may be concerned that the analysis presented in Figure 1 averages probability correct values across several different values of the stimuli on the current and the previous trial. For example, the central point in Figure 1 ($x = 0$) shows the average accuracy value across all repeated stimuli (i.e., for pairs of trials n, $n - 1$ where $S_n = S_{n-1} = j, j = 1, \ldots, N$). We also analyzed the data separately (conditional on the ordinal value of S_{n-1}), to check for a similar pattern of misfit between the RJM and the data for each value of S_{n-1}. Figure 2 shows this analysis, with solid lines for the predictions of the RJM and solid circles with dotted lines for the data. Each panel of Figure 2 shows accuracy for all 10 stimuli, conditional on just 1 prior stimulus (S_{n-1}). In every panel, the RJM predicts a large peak in accuracy when the current stimulus is near the previous stimulus and low accuracy when the current stimulus is far from the previous stimulus. Once again, the data do not support these predictions.

Can the Relative Judgment Model Be Rescued?

The predictions of the RJM with the given parameter values are clearly not supported by the data. However, it is possible that there are other parameter values that give suitable fits to Lacouture's (1997) data. In particular, the predicted accuracy of responding to extreme, and other, stimuli is influenced by the value of the criterion parameter *C*. The effects of this parameter are illustrated in Figure 3 (adapted from Stewart et al.'s, 2005, Figure 9). In the RJM, direct numerical values are assigned to stimuli, with the mental representations of the stimuli for Lacouture's experiments assumed to be centered on the numbers 1, 2, . . . , 10, as shown by the dotted vertical lines in Figure 3. Responses in the RJM are determined by setting criteria on this axis (forming a partition). The naive expectation may be to have the criteria placed halfway between the stimulus values, that is, at 1.5, 2.5, . . . , 9.5. This situation is illustrated by the bottom row of criteria (solid lines) in Figure 3, with $C = 0$. Stewart et al. (2005) set $C = 0.159$ in their fit of the assimilation and contrast effects in Lacouture's data; this value is in the top row of Figure 3. This parameter setting leads to slightly higher overall accuracy than when $C = 0$ because of the more widely spaced criteria.

The problem we identified above concerned the low accuracy predicted for extreme stimuli under certain conditions. We tried to remedy this problem by decreasing the *C* parameter, thus bringing the most extreme criteria toward the center and increasing predicted response accuracy to end stimuli. However, this adjustment had the undesired effect of decreasing the predicted response accuracy to near-to-extreme stimuli. We used standard optimization techniques[5] to identify a parameter set that improved the model's fit to Figures 1 and 2 while maintaining the model's predictions for other essential phenomena (including assimilation and contrast effects). We identified the following parameter set as providing the best fit we could find: $C = 0.0618$, $\sigma = 0.135$, $\lambda =$

[4] That code is available from Scott Brown's website (http://science-it.newcastle.edu.au/~sdb231/software/rjm/) and is written in the R language (but is brief and easy to translate into other languages). There is one minor detail of the simulations not mentioned in Stewart et al.'s (2005) article that is required to match their model predictions. The value of the stimulus presented *k* trials previously must be assumed to be the average stimulus value, where *k* is one larger than the number of α parameters. For the particular instantiation in this analysis, there were five α parameters, so the stimulus six trials prior to the current trial must be assumed to have ordinal value of 5.5 (the average value of the 10 stimuli). Omitting this detail (by simulating all possible combinations of the last six stimuli) leads to arguably more correct model predictions, but those predictions differ slightly from those presented in Stewart et al. (2005).

[5] As an ad hoc solution, we calculated the root-mean-square error between data and the RJM predictions for several different graphs, including those for assimilation, contrast, the bow effects in d' and accuracy, and the conditional accuracy graphs shown in Figure 2. We then used a standard simplex algorithm to minimize this error as a function of the parameter set for the RJM.

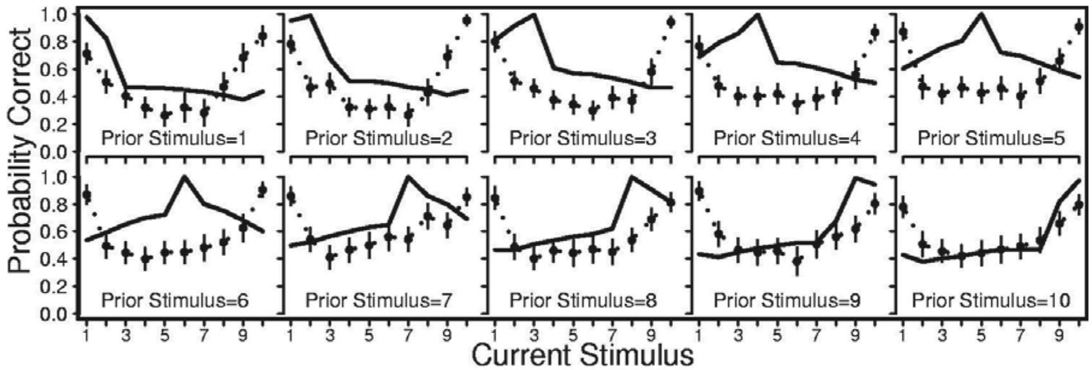

Figure 2. Probability of a correct response in data from Lacouture (1997; solid circles and dotted line) and in predictions from the relative judgment model (RJM; solid line). In each graph, the *x*-axis represents stimulus magnitude, and the separate panels show data conditional on the ordinal value of the previous stimulus, from 1 to 10. Error bars are ±2 standard errors.

0.877, $\alpha_1 = .173$, $\alpha_2 = .164$, $\alpha_3 = .134$, $\alpha_4 = .102$, $\alpha_5 = .0857$. Note that the *C* parameter we have estimated is about 60% smaller than the value used by Stewart et al. (2005). The other parameter values are all near to values used by Stewart et al., with some adjustment to compensate for the smaller value of *C*.

The effect of the much smaller *C* parameter was to move the response criteria toward the center, as shown in the middle row of Figure 3. The change was most dramatic for the extreme stimuli, leading to greater accuracy for those stimuli under the new parameter values. This change allowed the RJM to account for the W-shaped pattern of probability correct evident in Lacouture's (1997) data, which it previously could not (cf. Figure 1). Figure 4 (left panel) replicates Figure 1 but includes predictions of the RJM calculated using the new parameter values. The RJM now accounts for the probability correct data, at least qualitatively. There is some evidence of misfit: The RJM overestimates accuracy for repeated stimuli and underestimates some other accuracy values. However, the principle qualitative problem previously identified—that the predicted response accuracy was very low for extreme values—has been rectified. As already noted, these parameter values also retain the other important properties of the model, including the ability to accommodate assimilation and contrast phenomena.

By exploiting the RJM's criterion setting parameter, we were able to eliminate the erroneous prediction of a small probability of a correct response for an extreme stimulus preceded by the other extreme stimulus. However, a deeper analysis shows that our success was only partial. Thus far, the analysis has been only for correct responses, with the predictions improved as a result of changes in a response bias parameter, *C*. However, a complete analysis requires us to also analyze incorrect responses (errors). For such analysis, Luce, Nosofsky, Green, and Smith (1982) proposed a sensitivity measure, analogous to d' in signal detection theory, that is intended to separate bias effects (such as those because of *C* above) from discrimination (d') effects. Stewart et al. (2005, p. 896) admitted that the current version of their model underestimated such d' effects in various data, and now we show that it also does so in Lacouture's (1997) data.

For each pair of adjacent stimuli (i.e., 1 + 2, 2 + 3, . . . , 9 + 10) hit and false alarm rates are calculated by collapsing data across all larger and smaller response categories. Figure 4 (right panel) shows the choice data from Lacouture (1997) transformed using this measure. As has often been observed (e.g., by Luce et al., 1982), the bow effect evident in the raw accuracy data (left panel) is maintained in the sensitivity data (right panel). This can be interpreted as meaning that participants do not respond more accurately to extreme stimuli than to central stimuli simply because of response biases but that they are actually more *sensitive* at discriminating between extreme stimuli.[6] The transformed predictions from the RJM, on the other hand, show almost no bow effect in sensitivity. Indeed, the edges of the bow plot predicted by the RJM show a flattening.

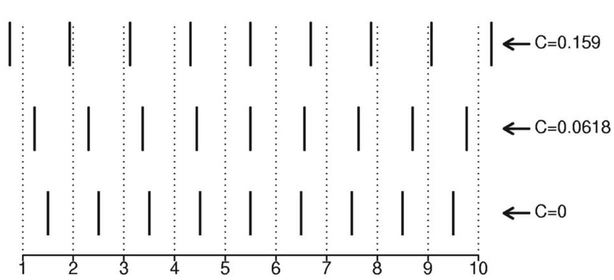

Figure 3. Criterion placement in the relative judgment model. As the *C* parameter increases (bottom to top) the criterion values move further from the center and from the midpoints between adjacent stimuli.

Discussion

Stewart et al.'s (2005) RJM successfully models complicated sequential phenomena such as assimilation and contrast. As summarized above, this success is based on three major assumptions: first, that responses to the stimulus presented on the current trial are made relative to a context determined by comparisons of that stimulus with

[6] It is possible that one may question the theoretical interpretation of the above d' measure as separating bias and sensitivity effect. Nevertheless, the measure summarizes many aspects of the data, and a satisfactory theory should fit data even when the data and the theory's predictions have been transformed by such a summary measure.

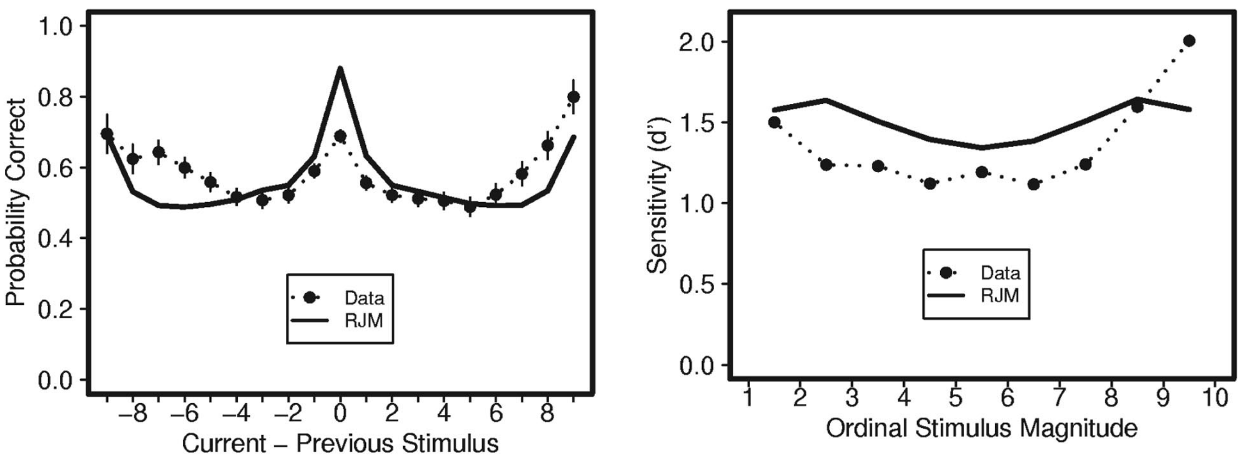

Figure 4. The left panel replicates Figure 1, showing response accuracy conditional on the difference between successive stimuli but using model predictions with new parameter estimates (see text). The right panel shows the choice data transformed to sensitivity measurements (d'). RJM = relative judgment model.

previous stimuli and responses and feedback (if given); second, that the set of available responses is constrained by the relative ordering of the stimulus on the current and the previous trial; and third, that the variability of the responses to the current stimulus is an increasing function of the number of available responses. Together, these three assumptions, with the parameter values used by Stewart et al., lead to the prediction that accuracy will be very low when a large stimulus is followed by a small stimulus, or vice versa. Data do not confirm this prediction. For instance, Lacouture's (1997) data show the *largest* accuracy for large-followed-by-small or small-followed-by-large stimuli (presumably because end stimuli are well identified). By adjusting the response criteria through the *C* parameter, we were able to identify a set of parameter values that allowed the RJM to fit the qualitative patterns of the correct responses. However, this account of the data did not hold up when both correct and error responding were considered simultaneously with the d' sensitivity measure. This is a further example of the RJM's failure to well-fit results plotted in terms of such a sensitivity measure, a weakness of the model that the authors acknowledge in their article (Stewart et al., 2005, p. 896).

Although our analyses point to a problem in the RJM, it is not clear at what point the problem arises. It could be due to the fact that the current version of the model does not include any detailed representations of the stimuli that would allow naturally for discrimination effects. Alternatively, one could argue that the most basic implication of our results is that the idea of relative judgment must be tempered: Perhaps observers really *do* keep some kind of absolute memory for edge stimuli. Alternatively, one could argue that the problem begins in the early processing stages of the RJM, with the assumption that the observer can very accurately know whether the current stimulus is smaller than, equal to, or larger than the prior stimulus. This assumption is almost certainly not true in real data. For example, Nosofsky (1983) showed that the above kind of comparative (discrimination) task—in the context of absolute identification—is not performed perfectly accurately (e.g., p. 106, where sensitivity in the discrimination condition falls to around $d' = 1.1$ for some stimulus pairs).

The limitations of Stewart et al.'s (2005) fit of the RJM to Lacouture's (1997) data were masked in their article because they fit averaged data.[7] With suitable parameter values, the RJM gives a reasonable account of the conditional response accuracy data, although the fit is seen to be unsatisfactory when the data are displayed via the discrimination (d') measure.

[7] Of course, some averaging across trials is unavoidable. For instance, estimates of probability correct on the current trial n, on the basis of, say, specific values of the stimuli on trial $n - 1$ *and* on trial $n - 2$, will likely be quite unreliable.

References

Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response-time analysis. *Psychological Research, 60,* 121–133.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics, 32,* 397–408.

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

Mori, S., & Ward, L. M. (1995). Pure feedback effects in absolute identification. *Perception & Psychophysics, 57,* 1065–1079.

Nosofsky, R. M. (1983). Shifts of attention in the identification and discrimination of intensity. *Perception & Psychophysics, 33,* 103–112.

Rouder, J. N., Morey, R. D., Cowan, N., & Pfaltz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review, 11,* 938–944.

Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review, 101,* 357–361.

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

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9,* 73–78.

Received March 8, 2006
Revision received June 26, 2006
Accepted June 27, 2006 ■