Psychological Review
2012, Vol. 119, No. 1, 216–220

0033-295X/12/$12.00 DOI: 10.1037/a0025824

COMMENT

The Law of Categorical Judgment (Corrected) Extended: A Note on Rosner and Kochanski (2009)

Karl Christoph Klauer and David Kellen
Albert-Ludwigs-Universität Freiburg

Rosner and Kochanski (2009) noticed an inconsistency in the mathematical statement of the Law of Categorical Judgment and derived "the valid equation, the Law of Categorical Judgment (Corrected)" (p. 125). The purpose of this comment is to point out that the law can be corrected in many different ways, leading to substantially different equations. The different versions have different consequences for the predicted distributions of the responses and, hence, for fitting real data. Some of these consequences are unexpected and sometimes undesirable. Researchers should be aware of the different possibilities as they may lead to pronouncedly different accounts of given data.

Keywords: Law of Categorical Judgment, rating data, signal detection theory, mathematical models

Rosner and Kochanski (2009), henceforth referred to as RK, considered the famous Law of Categorical Judgment (McNicol, 1972; Torgerson, 1958; Wickelgren, 1968). It describes responses of a rating experiment in which stimuli are presented from N classes of stimuli, and the subject responds with one of $M + 1$ possible ordered responses R_i.

As stated by RK, the Law of Categorical Judgment supposes that any stimulus S_h from stimulus class h projects a Gaussian signal density $\varphi(x;\, \mu_{Sh}, \sigma_{Sh})$ of subjective impressions on at least an interval-scaled psychological decision axis ν.[1] The observer furthermore defines the $M + 1$ responses by placing M criteria at different loci on ν on each trial. Criterion positions c_i are samples from underlying densities $\varphi(x;\, \mu_{Ci}, \sigma_{Ci})$; see Figure 1. The variability in criterion positions that emerges from sampling when $\sigma_{Ci} > 0$ is sometimes termed *criterion noise* (e.g., Benjamin, Diaz, & Wee, 2009; Mueller & Weidemann, 2008).

On any trial, the observer supposedly has one signal sample with value s on ν and M indexed criterion samples with values c_i, $i = 1, \ldots, M$ (Rosner & Kochanski, 2009, p. 117). Only the experimenter knows the stimulus class h from which the stimulus value s was sampled. Only the observer knows the ordinal index i of each c_i. Note that even if $i < j$, it may be that $c_i > c_j$ in the situation studied by RK.

As explained by RK, the long accepted mathematical expression of the Law of Categorical Judgment is the distance between the means of two Gaussian distributions, one for the signal class and one for the ith criterion

$$z_{hi} = (\mu_{Ci} - \mu_{Sh})/(\sigma^2_{Ci} + \sigma^2_{Sh})^{1/2}$$

where z_{hi} is a unit normal deviate. The theoretical cumulative probability that S_h is rated between 1 and i, $P(R \leq i|S_h)$, equals $\Phi(z_{hi};\, 0, 1)$, where Φ is the Gaussian distribution function. Following RK's exposition, this assumes that signal and criterion values are independent random variables.

RK noticed that the model as stated is inconsistent. Given unequal criterion variances, it can predict negative values for the probability of response i, $P(R = i|S_h) = P(R \leq i|S_h) - P(R \leq i - 1|S_h)$, when $2 \leq i \leq M$. RK went on to derive a model that does not contain the inconsistency and reduces to the signal detection theory (SDT) model for rating data defined by the absence of criterion variability (i.e., with $\sigma^2_{Ci} = 0$ for all i). As stated by RK, "we have derived the valid equation, the Law of Categorical Judgment (Corrected)" (p. 125).

The purpose of this comment is to point out that the Law of Categorical Judgment can be corrected in different ways. In fact, there is an infinite number of ways in which the Law of Categorical Judgment can be consistently defined in the presence of criterion noise. The different versions have different consequences for the predicted response distributions and hence for fitting data. The consequences can be unexpected and sometimes undesirable.

The crux of the problem is how the decision rule is chosen. The decision rule maps an ensemble of criterion samples c_i and signal sample s in a given trial on an observable response. The rule chosen by RK is the following (Rule 1): "Under the Law of Categorical Judgment, the observer uses the rating scale decision rule: *Calculate the M differences $c_i - s$. If $c_i - s$ is the smallest positive difference, make response i; if all $c_i - s$ are negative, make response $M + 1$*" (p. 117).

In the literature given by RK as definitive for the Law of Categorical Judgment (i.e., McNicol, 1972; Torgerson, 1958;

Karl Christoph Klauer and David Kellen, Institut für Psychologie, Albert-Ludwigs-Universität Freiburg.

The research reported in this paper was supported by grant Kl 614/31-1 from the Deutsche Forschungsgemeinschaft to the first author.

Correspondence concerning this article should be addressed to Karl Christoph. Klauer at the Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Freiburg D-79085, Germany. E-mail: christoph.klauer@psychologie.uni-freiburg.de

[1] For ease of comparison, we use the same notation as adopted by RK.

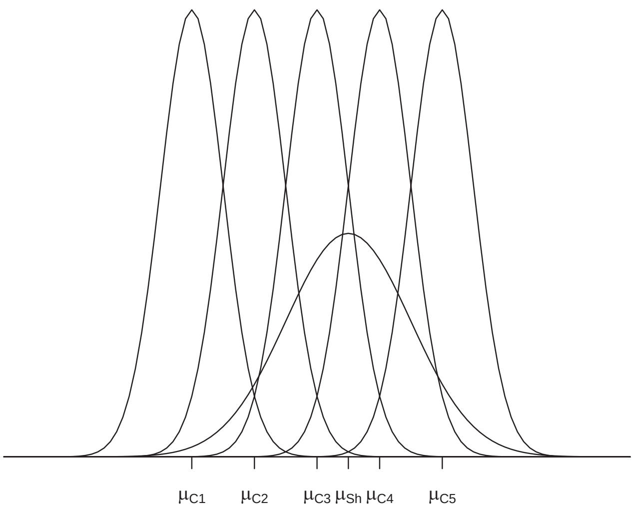

Figure 1. Distributions of criteria and signal on the decision axis.

Wickelgren, 1968), we found no statement of this rating scale decision rule. In fact, the problem of determining the appropriate response is complex in the situation studied by RK, that is, if criteria are allowed to change ordinal positions, a possibility that finds support in previous research (Treisman & Faulkner, 1985). The decision rule proposed by RK is a consistent one, but there are many others that could be adopted. Two obvious options are Rules 2 and 3, in order:

- *Calculate the M differences $s - c_i$. If $s - c_i$ is the smallest positive difference, make response $i + 1$; if all $s - c_i$ are negative, make response* 1; and
- *Calculate the M absolute differences $|s - c_i|$. If $|s - c_i|$ is the smallest absolute difference, make response i, if $c_i - s$ is positive; make response $i + 1$, if $c_i - s$ is negative.*

Does It Matter?

What are the consequences of adopting either of these decision rules? Consider first the rule elaborated on by RK. One, perhaps unexpected, consequence is that it produces asymmetrical probability distributions of rating responses in symmetrical situations. Assume that there are six responses and five criteria c_1 to c_5 with means, in order, −2, −1, 0, 1, and 2 on ν. Assume furthermore that the signal mean μ_{Sh} is 0 and that signal and criterion variances equal 1. Thus, we have a situation in which the signal is centered on the same value as the middle criterion c_3 and the criterion means are spaced out symmetrically to the right and to the left moving outward from that middle point.

Most researchers, at least those in the SDT field, would probably expect that these parameter values would translate into a predicted distribution of responses that is symmetrical around the scale midpoint 3.5, so that $P(R = 1) = P(R = 6)$, $P(R = 2) = P(R = 5)$, and $P(R = 3) = P(R = 4)$. Figure 2 shows the response probabilities predicted under the Law of Categorical Judgment (Corrected); see Figure 2, white bars. As can be seen, the expectation is not fulfilled, and the asymmetry is quite pronounced; compare, e.g., $P(R = 2)$ and $P(R = 5)$. Figure 2 also shows the response probabilities for a signal distribution with $\sigma_{Sh} = 2$ (see Figure 2, black bars). As can be seen, the asymmetry can take different shapes—for example, for $\sigma_{Sh} = 1$, $P(R = 1|S_h) > P(R = 6|S_h)$, and vice versa for $\sigma_{Sh} = 2$.

What is the cause of the asymmetry? The cause is the asymmetry built into the decision rule. It focuses on criteria to the right of the signal value (by considering only positive differences $c_i - s$). $P(R = M + 1)$ therefore corresponds to the probability that s is larger than all c_i irrespective of the ordering of the criteria. Due to the symmetrically placed underlying parameters, this will be the same as the probability that s is smaller than all c_i. But $P(R = 1)$ is not given by this probability; instead, it corresponds to the sum of the probabilities of all possible arrangements of s and c_i for which $c_1 - s$ is the smallest positive difference. This means that the probabilities $P(R = 1)$ and $P(R = M + 1)$ will differ in general.

The second decision rule implies a similar asymmetry, but its shape will be the mirror image of that of RK's decision rule. There may be good reasons for having a model exhibiting asymmetry. For example, the decision scale ν may be an inherently asymmetrical scale such as a strength-of-evidence dimension with larger values indicating more evidence for the presence of a signal without a natural midpoint. In such cases, it may make sense to treat small and large values on that axis differently. If so, we would probably still want to justify why the asymmetry should have the specific shape implied by RK's decision rule and not, for example, the one implied by the second decision rule. In the absence of good reasons to expect a specific form of asymmetry, the choice of a symmetrical model may be a reasonable default option.

In other cases, it makes little sense a priori to have an asymmetrical model. For example, in source memory, a rating scale might be used with six values, three to the left, encoding in order, high, middle, and low confidence in a Source A attribution, and three to the right, encoding in order, low, middle, and high con-

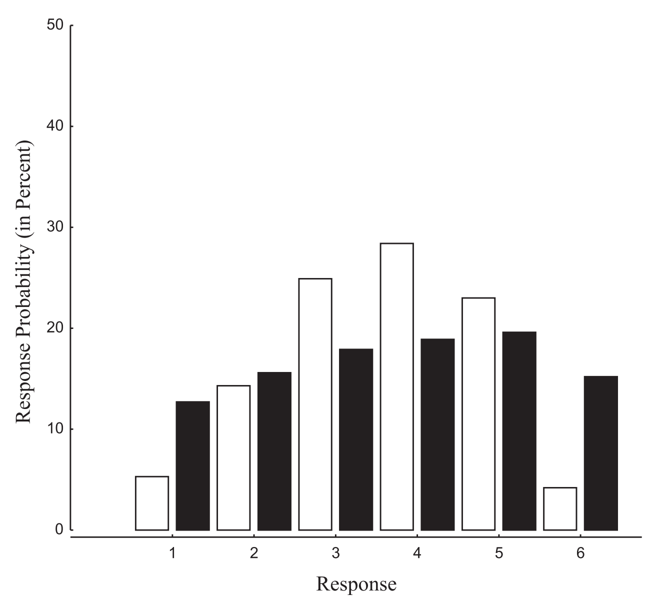

Figure 2. Example distributions of responses for symmetrical underlying parameters under Rosner and Kochanski's (2009) Law of Categorical Judgment (Corrected). White and black bars each define one distribution; the two distributions differ in the assumed signal variance.

fidence in a Source B attribution (e.g., Yonelinas, 1999). Most researchers would probably expect that a model would predict symmetrical distributions of responses in a situation with symmetrically placed underlying parameters.

Another example in which asymmetry is problematic is the two-alternative forced choice (2AFC) rating task (e.g., Jang, Wixted, & Huber, 2009). In the 2AFC rating task, pairs of items, S_1 and S_2, are presented, and confidence ratings are often obtained for the spatial (e.g., S_1 is left vs. right) or temporal (e.g., S_1 was first vs. S_2 was first) positions of the signals. In such cases, the coding of the responses is assumed to be arbitrary, as no consequences in terms of the model account are expected to hinge on it (Wickens, 2002, p. 101).

One consequence of the asymmetry is, however, that it would make a difference in model fitting whether Rule 1 or Rule 2 is chosen or, equivalently, whether responses are entered as recorded or recoded by mirroring responses at the scale midpoint (i.e., replacing response r by $7 - r$ in the previous example). The expectation would probably be that model fit is the same for both data sets and that parameter estimates are essentially the same (i.e., mean criterion and signal placements from both data sets should essentially be mirror images of each other on the decision scale ν and variances equal, after correcting for possibly different fixings of zero point and scale). In fact, neither of the expectations is true.

We multiplied the probabilities in Figure 2 by 1,000 and entered these as data in fitting RK's model allowing for unequal signal variances with criterion variances restricted to be equal (this is the restricted model recommended by RK for fitting data from just two signal distributions). We did this twice, once for the original data and once for recoded data with responses mirrored at the midpoint of the rating scale.

Table 1 (see rows labeled Example 1) shows the results of the model fitting. The first row gives parameter estimates and goodness-of-fit value G^2 for the model implied by RK's rule (i.e., Rule 1); the second row shows the results for the recoded data (with criterion and signal means mirrored, for ease of comparison, at the mean of the first signal, which was fixed to zero, defining the zero point of the decision scale for both analyses). Note that the results shown in the second row thereby also correspond to fitting the model implied by Rule 2 to the original (nonrecoded) data.[2] As can be seen, the fitting recovers the underlying parameters and produces a perfect fit for the original data using RK's model. After recoding, model fit is worse, and criterion means as well as signal and criterion variances change noticeably. For example, criterion noise is underestimated; in fact, forcing the criterion variance to be equal to the variance of S_1 (as is the case in the underlying parameters) leads to a significant deterioration in model fit, $G^2(1) = 11.87$, $p < .01$.

In this example, both signals are centered on zero, which is theoretically possible but unusual. For this reason, we analyzed a second example with the sole difference being that the second signal's mean was set to 1, $\mu_{S2} = 1$, so that S_1 and S_2 are displaced by one unit on the underlying scale. Table 1 (see rows labeled Example 2) shows the results of the model fitting. As can be seen, the results are analogous to those reported for the first example.

The Law of Categorical Judgment (Symmetrically Corrected)

Rule 3 defines a symmetrical model. Because a symmetrical model may be preferred in the analysis of rating data for the previously listed reasons, we give the equations for the induced distribution of responses, which might be termed the Law of Categorical Judgment (Symmetrically Corrected):[3]

$$P(R = 1|S_h) = \int \varphi(s; \mu_{Sh}, \sigma_{Sh}) \int_{s}^{\infty} \varphi(c_1; \mu_{C1}, \sigma_{C1}) \prod_{j \neq 1} \left(1 - \int_{2s-c_1}^{c_1} \varphi(c_j; \mu_{Cj}, \sigma_{Cj}) dc_j\right) dc_1 ds,$$

$$P(R = i|S_h) = \int \varphi(s; \mu_{Sh}, \sigma_{Sh}) \int_{s}^{\infty} \varphi(c_i; \mu_{Ci}, \sigma_{Ci}) \prod_{j \neq i} \left(1 - \int_{2s-c_i}^{c_i} \varphi(c_j; \mu_{Cj}, \sigma_{Cj}) dc_j\right) dc_i ds$$
$$+ \int \varphi(s; \mu_{Sh}, \sigma_{Sh}) \int_{-\infty}^{s} \varphi(c_{i-1}; \mu_{C(i-1)}, \sigma_{C(i-1)}) \prod_{j \neq i-1} \left(1 - \int_{c_{i-1}}^{2s-c_{i-1}} \varphi(c_j; \mu_{Cj}, \sigma_{Cj}) dc_j\right) dc_{i-1}\, ds,$$

$$\text{for } 2 \leq i \leq M, \text{ and}$$

$$P(R = M + 1|S_h) = \int \varphi(s; \mu_{Sh}, \sigma_{Sh}) \int_{-\infty}^{s} \varphi(c_M; \mu_{CM}, \sigma_{CM}) \prod_{j \neq M} \left(1 - \int_{c_M}^{2s-c_M} \varphi(c_j; \mu_{Cj}, \sigma_{Cj}) dc_j\right) dc_M ds$$

Note that these decision rules imply the same responses as the traditional SDT model for rating data when $c_1 < c_2 < \ldots < c_M$; hence, the SDT model is the limiting case for all of the previously shown decision rules when there is no criterion variance.

Note also that the decision rules by no means exhaust the space of reasonable decision rules that could be chosen. For example,

[2] We constrained the criterion means μ_{Ci} to be ordered as $\mu_{C1} \leq \mu_{C2} \leq \ldots \leq \mu_{CM}$ because a switch in the order of the means would presumably make little sense. If this restriction is removed, the analysis using Rule 1 is unchanged, but the best-fitting parameter values using Rule 2 exhibit such switches.

[3] The response probabilities are the sum of two terms for $2 \leq i \leq M$. The first term gives the probability that $|c_i - s|$ is the smallest difference and $c_i \geq s$. The product over $j \neq i$ in that term gives the probability that all other $|c_j - s|$ are larger, given c_i and s with $c_i \geq s$; this means that either $c_j - s > c_i - s$ or $s - c_j > c_i - s$ (i.e., $c_j > c_i$ or $c_j < 2s - c_i$). The second term gives the probability that the $|c_{i-1} - s|$ is the smallest difference and $c_{i-1} \leq s$.

Table 1
Two Examples of Fitting Rosner and Kochanski's Model to Two Data Sets Differing Only in Mirrored Coding of Responses: Maximum Likelihood Estimates and Goodness-of-Fit Statistic G^2 (df = 2)

Data/rule	μ_{S1}	σ_{S1}	μ_{S2}	σ_{S2}	μ_{C1}	μ_{C2}	μ_{C3}	μ_{C4}	μ_{C5}	σ_C	G^2	p
Example 1												
Original/Rule 1	0.00[a]	1.00[a]	0.00	2.00	−2.00	−1.00	0.00	1.00	2.00	1.00	0.00	1.00
Recoded[b]/Rule 2	0.00[a]	1.00[a]	0.06	1.59	−1.74	−0.92	−0.18	0.60	1.76	0.33	3.79	.15
Example 2												
Original/Rule 1	0.00[a]	1.00[a]	1.00	2.00	−2.00	−1.00	0.00	1.00	2.00	1.00	0.00	1.00
Recoded[b]/Rule 2	0.00[a]	1.00[a]	0.84	1.69	−1.73	−1.10	−0.16	0.55	1.80	0.42	3.04	.22

[a] Fixed a priori to determine zero point and scale. [b] Estimates of signal and criterion means are mirrored at $\mu_{S1} = 0$.

one could compare s and c_i in a fixed order, starting, for example, with c_1, followed by c_2, and so forth, and respond "i" as soon as a c_i is found with $s \leq c_i$ (and respond "$M + 1$" if none such exists). Mueller and Weidemann (2008) similarly proposed to consider first whether s falls left or right of a middle criterion (more precisely, a criterion with index in the middle, i.e., $c_{(M+1)/2}$ assuming M is an odd number) and to then move outwards (in terms of criterion indices) from the center in that (left or right) direction until a criterion is found that is further away from the midpoint than s, and so forth.[4] Finally, observers might adopt probabilistic mixtures of different such rules.

Discussion

An interesting question concerns the Law of Categorical Judgment (Uncorrected) for the case of equal criterion variances. As explained by RK, positional interchanges of criteria can still happen even with equal criterion variances, but the rate of exchange for each pair of adjacent criteria would be sufficiently balanced to prevent overt negative probabilities (Rosner & Kochanski, 2009, p. 118). In this case, the law does define a probability distribution of the responses in terms of parameters labeled just as the parameters assumed to underlie the signal and criterion distributions, that is (for one signal class h) as $\theta = (\mu_{Sh}, \sigma_{Sh}, \mu_{C1}, \ldots, \mu_{CM}, \sigma_C)$. Because it defines a probability distribution, it is possible to fit the original Law of Categorical Judgment with the restriction of equal criterion variances (which can be absorbed into the signal variances to have an identified model) given data for trials with at least two signal classes. A much noted article by Benjamin et al. (2009) gives an example.

The uncorrected law describes the probability distribution of responses if criterion noise takes the form of a common random additive shift that is normally distributed and added to the otherwise fixed criteria. This implies perfect intercorrelations of all criteria along with equal criterion variances and no interchanges in ordinal positions of criteria. Hence, the parameters of the uncorrected law can be interpreted as the means and variances of the underlying signal and criterion distributions in this situation. But do they describe these quantities of interest when criteria are allowed to interchange ordinal positions?

The Law of Categorical Judgment (Corrected) does not reduce to the uncorrected law in this situation even if the symmetrical Rule 3 is chosen (it cannot reduce to the uncorrected law if one of the asymmetrical Rules 1 or 2 is chosen because the uncorrected Law of Categorical Judgment is a symmetrical model). This follows from the fact that (for $s < c_1$)

$$\prod_{j \neq 1}\left(1 - \int_{2s-c_1}^{c_1} \varphi(c_j;\, \mu_{Cj}, \sigma_{Cj})dc_j\right) < 1,$$

as soon as there is criterion variance. Hence

$$P(R = 1) < \int \varphi(s;\, \mu_{Sh}, \sigma_{Sh}) \int_s^{\infty} \varphi(c_1;\, \mu_{C1}, \sigma_{C1})dc_1 ds = \Phi(z_{h1};\, 0{,}1).$$

But $P(R = 1)$ should be equal to $\Phi(z_{h1};\, 0, 1)$ according to the uncorrected law. It follows that the parameter estimates obtained by fitting the uncorrected law (with the restriction of equal criterion variances) will not be consistent estimates of the parameters underlying the corrected law, even if criterion variances are equal.

But what is the relation of the parameters of the uncorrected law on the one hand and the parameters assumed to underlie the signal and criterion distributions on the other hand in the case of equal criterion variances (and less than perfectly intercorrelated criteria)? A link between the two sets of parameters would be established by a decision rule that leads to the Law of Categorical Judgment (Uncorrected) as describing overt responses at least in the case of equal criterion variances, with θ being the parameters of the underlying signal and criterion distributions. We believe that no such decision rule exists, but have not found a proof of its nonexistence. In consequence, the question of how to interpret the outcome of fitting the uncorrected law with equal criterion variances is open.

If such a rule exists, it would probably have the most legitimate claim to being the appropriate correction of the Law of Categorical

[4] Mueller and Weidemann (2008) developed this for binary classifications and even and odd numbers of responses and criteria, respectively, but it is straightforward to extend the idea to the much more general situation encompassed by the Law of Categorical Judgment. This leads to another symmetrical decision rule. Yet another symmetrical decision rule in this spirit might first determine the location of the signal with respect to a middle criterion and then move inward from the criterion with the most extreme index on the side of the signal rather than outward from the middle criterion as in Mueller and Weidemann (2008). One sometimes undesirable feature of these symmetrical decision rules is that they treat the middle criterion $c_{(M+1)/2}$ (assuming M is odd) differently from the other criteria, which would be reflected in the estimates of their means and variances.

Judgment. As it stands, the law can be corrected in different ways with different consequences. One perhaps unexpected and sometimes undesirable feature of the version proposed by RK is a built-in asymmetry. It is, however, possible to correct the law in RK's spirit while preserving the symmetry of the original Law of Categorical Judgment. Researchers should be aware of the different possibilities as they may lead to pronouncedly different accounts of given rating data as exemplified above.

One exception is the case of equal criterion variances with perfect intercorrelations of criteria as already mentioned. In this case, criterion noise takes the form of an additive random shift added to all criteria as a block and interchanges in ordinal positions of the criteria do not occur. This in turn implies that all of the previously described decision rules generate the same predicted distribution of responses described by the uncorrected Law of Categorical Judgment.

This set of assumptions is perhaps not very plausible, but the assumption of perfectly independent criteria adopted by RK and others might similarly be questioned. If less than perfect criterion intercorrelations are permitted, positional interchanges can occur, and the analyses would therefore still depend upon the particular decision rule adopted. The formulae would, however, have to be built on a multivariate normal distribution of the criteria with additional parameters introduced in the form of the variance–covariance matrix Σ of that distribution. Relaxing the independence assumption would therefore introduce new parameters and make the formulae more complex, but over and above the added complexity, the same issues arise as for the case of independent criteria.

Analyses of rating data are thereby premised on the choice of a decision rule if criteria are allowed to change ordinal positions, and it may be prudent to check the major conclusions obtained from such analyses for robustness under different choices of decision rule. On the other hand, these observations can also be turned around: Because different decision rules induce different models of the rating data, comparing the different models' fit to real data might shed some light on the decision rule actually preferred by respondents.

References

Benjamin, A. S., Diaz, M. L., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review, 116,* 84–115.

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal detection models of yes/no and two-alternative-forced-choice recognition memory. *Journal of Experimental Psychology: General, 138,* 291–306.

McNicol, D. (1972). *A primer of signal detection theory.* Mahwah, NJ: Erlbaum.

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review, 15,* 465–494.

Rosner, B. S., & Kochanski, G. (2009). The Law of Categorical Judgment (Corrected) and the interpretation of changes in psychophysical performance. *Psychological Review, 116,* 116–128.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York, NY: Wiley.

Treisman, M., & Faulkner, A. (1985). Can decision criteria interchange locations? Some positive evidence. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 187–208.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology, 5,* 102–122.

Wickens, T. D. (2002). *Elementary signal detection theory.* New York, NY: Oxford University Press.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415–1434.

Received April 17, 2011
Revision received July 25, 2011
Accepted August 29, 2011 ■