Psychonomic Bulletin & Review
2007, 14 (5), 858-865

A direct test of the unequal-variance signal detection model of recognition memory

Laura Mickes, John T. Wixted, and Peter E. Wais
University of California, San Diego, La Jolla, California

Analyses of the receiver operating characteristic (ROC) almost invariably suggest that, on a recognition memory test, the standard deviation of memory strengths associated with the lures (σ_{lure}) is smaller than that of the targets (σ_{target}). Often, $\sigma_{\text{lure}}/\sigma_{\text{target}} \approx 0.80$. However, that conclusion is based on a model that assumes that the memory strength distributions are Gaussian in form. In two experiments, we investigated this issue in a more direct way by asking subjects to simply rate the memory strengths of targets and lures using a 20-point or a 99-point strength scale. The results showed that the standard deviation of the ratings made to the targets (s_{target}) was, indeed, larger than the standard deviation of the ratings made to the lures (s_{lure}). Moreover, across subjects, the ratio $s_{\text{lure}}/s_{\text{target}}$ correlated highly with the estimate of $\sigma_{\text{lure}}/\sigma_{\text{target}}$ obtained from ROC analysis, and both estimates were, on average, approximately equal to 0.80.

Signal detection theory has long been a prominent theoretical framework for understanding how subjects make decisions on recognition memory tasks. The textbook version of the theory involves two equal-variance Gaussian distributions and a decision criterion placed somewhere along the memory strength axis. One distribution represents the memory strengths of the lures, and it has a low average value. The other distribution represents the memory strengths of the targets, and it has a higher average value. Any test item that generates a memory strength exceeding the criterion is declared to be old, otherwise it is declared to be new (as illustrated in the upper panel of Figure 1). Although the aesthetically appealing equal-variance version of the model is often used to illustrate signal detection theory, analyses of the empirical receiver operating characteristic (ROC) almost always imply an unequal-variance model in which the standard deviation of the target distribution exceeds that of the lure distribution (Egan, 1958, 1975; Ratcliff, Shue, & Gronlund 1992), as illustrated in the lower panel of Figure 1.

An ROC is simply a plot of the hit rate (HR) versus the false alarm rate (FAR) for different levels of bias. A typical ROC is obtained by asking subjects to supply confidence ratings for their recognition memory decisions, often on a 6-point scale. Signal detection theory predicts that the ROC will be curvilinear in probability space (HR vs. FAR) and linear in *z*-space (*z*-HR vs. *z*-FAR), and it holds that the slope of the *z*-ROC provides an estimate of the ratio of the standard deviation of the lure distribution to the standard deviation of the target distribution ($\sigma_{\text{lure}}/\sigma_{\text{target}}$). If an equal-variance model applies (as in the upper panel of Figure 1), then the slope should be 1.0. But if the standard deviation of the target distribution exceeds that of the lure distribution (as in the lower panel of Figure 1), then the slope of the *z*-ROC should be less than 1.0.

Previous reviews of the ROC literature indicate that *z*-ROCs are well characterized by a straight line and that the slope of the best-fitting line is, on average, approximately 0.80 (Glanzer, Kim, Hilford, & Adams, 1999; Ratcliff et al., 1992). Thus, according to the signal detection account, the standard deviation of the target distribution is often about 1.25 (i.e., 1/0.80) times that of the lure distribution. Findings like these explain why the unequal variance model shown in the lower panel of Figure 1 is regarded by some as the standard model of decision-making on a recognition memory task. Others, however, find the model to be less compelling. For example, the majority of investigations into the neuroanatomical basis of recognition memory either implicitly or explicitly reject this way of thinking (Wixted, 2007). If signal detection theory does provide an accurate model of decision-making, then those investigations could be led astray by the alternative decision-making models they embrace.

One issue that bears on the validity of the detection account is its suggestion that the standard deviation of the target distribution is greater than that of the lure distribution. That conclusion is based on an analysis that assumes that the underlying distributions of memory strength are Gaussian in form. Although ROC data are well fit by a Gaussian model, it has long been known that other distributions—ones that are quite unlike the Gaussian—also fit ROC data well. Instead of relying on ROC analysis, a more direct test of the unequal-variance idea would be to simply ask subjects to rate the memory strengths of targets and lures using a fine-grained scale (e.g., 1–99). The mean and standard deviation of the ratings for the targets (m_{target}

J. T. Wixted, jwixted@ucsd.edu

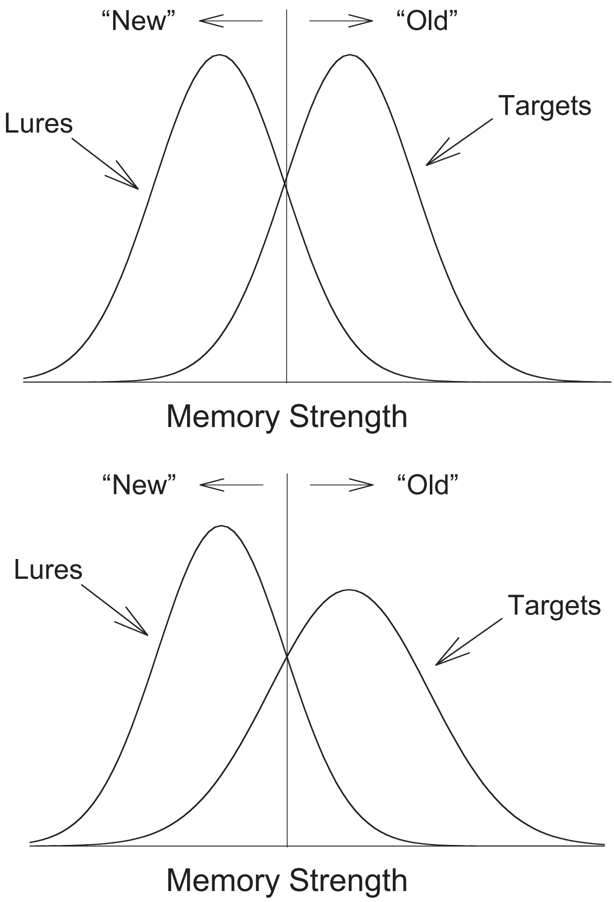

Figure 1. Equal-variance (upper panel) and unequal-variance (lower panel) signal detection models of recognition memory.

and s_{target}, respectively) could then be directly computed and then compared to the mean and standard deviation of the ratings for the lures (m_{lure} and s_{lure}, respectively). Although the mean rating for the targets would undoubtedly be greater than the mean rating for the lures, would the standard deviation of the target ratings be greater as well? And, if so, would the ratio of the standard deviation of the lure ratings to the standard deviation of the target ratings be approximately 0.80, as suggested by ROC analysis? These are the questions we set out to address.

EXPERIMENT 1

In the first experiment, subjects were presented with a list of 150 words to memorize, after which they completed a recognition memory test that involved those 150 targets randomly intermixed with 150 lures. For each test item, the subject was asked to rate the strength of their memory for that item on a 1–20 scale.

Method

Subjects. Fourteen undergraduates from University of California, San Diego, participated for lower-division psychology course credit.

Materials and Design. The word pool used consisted of 705 three-to-seven letter words taken from the MRC Psycholinguistic Database (Coltheart, 1981), of which 300 words were randomly selected for testing (150 of which were randomly selected to be targets, and the remainder were lures). Instructions and stimuli were displayed for each subject on an NEC MultiSync LCD1760NX monitor, and powered by a Dell Dimension 4550 computer. Stimuli were presented using an E-Prime program (www.pstnet.com; Psychology Software Tools).

Procedure. Subjects signed a consent form, were read instructions, studied the 150 targets, and completed a recognition test in which the 150 targets were randomly intermixed with the 150 lures. Each word was presented for 2 sec during study. During testing, subjects indicated whether or not the word was on the presented list by pressing a key; then they indicated the strength of their memory for that word by entering a number on the keypad ranging from 1 to 20, with 1 meaning that the word was definitely not on the list and 20 meaning that the word was definitely on the list. These instructions were given verbally prior to list presentation and appeared again on the screen after the list was presented. In addition, the verbal instructions asked subjects to be cautious about using the endpoints of 1 and 20. They were instructed to use those values only when they were 100% certain, one way or the other, as they might be if their own name was used as a test item.

Results

Subjects generally distributed their responses over the full range of the scale. The upper panel of Figure 2 shows the frequency distribution for targets and lures pooled over subjects. The lure distribution appears to be somewhat truncated on the left (as if some lures would have received lower ratings, if possible), and the target distribution appears even more truncated on the right (as if some targets would have received higher ratings), but the figure illustrates the central assumption of signal detection theory: The distributions of memory strengths for the targets and lures overlap, with the mean of the target distribution being higher than that of the lure distribution. The lower panel of Figure 2 shows decision accuracy for each rating. A rating in the range of 1 through 10 was scored as a correct response to lures (and an incorrect response to targets), whereas the reverse was true for ratings in the range of 11 through 20. In accordance with the predictions of signal detection theory, accuracy varies continuously as the distance from the indifference point increases.

As shown in Table 1, most subjects were relatively unbiased in their use of the rating scale such that their ratings for all items averaged together (m_{overall}) were close to the midpoint of the scale (10.5), with the overall mean being 10.77. However, Subject 11 was an exception. That subject's average rating across targets and lures was 14.2, which is 2.40 standard deviations above the mean. Indeed, even for lures, this subject's mean rating exceeded 10. This is an important consideration, because if a subject's ratings are biased toward one end of the scale (as this subject's ratings clearly are), the ratings for one class of items will be more compressed than the ratings for the other class. Except where noted, this subject was excluded from the main analysis.

Table 1 also shows the means and standard deviations for the ratings made to the targets (m_{target} and s_{target}, respectively) and to the lures (m_{lure} and s_{lure}) for each subject. Across all subjects (excluding Subject 11), the mean rating for the targets was 12.98, and the mean rating for the

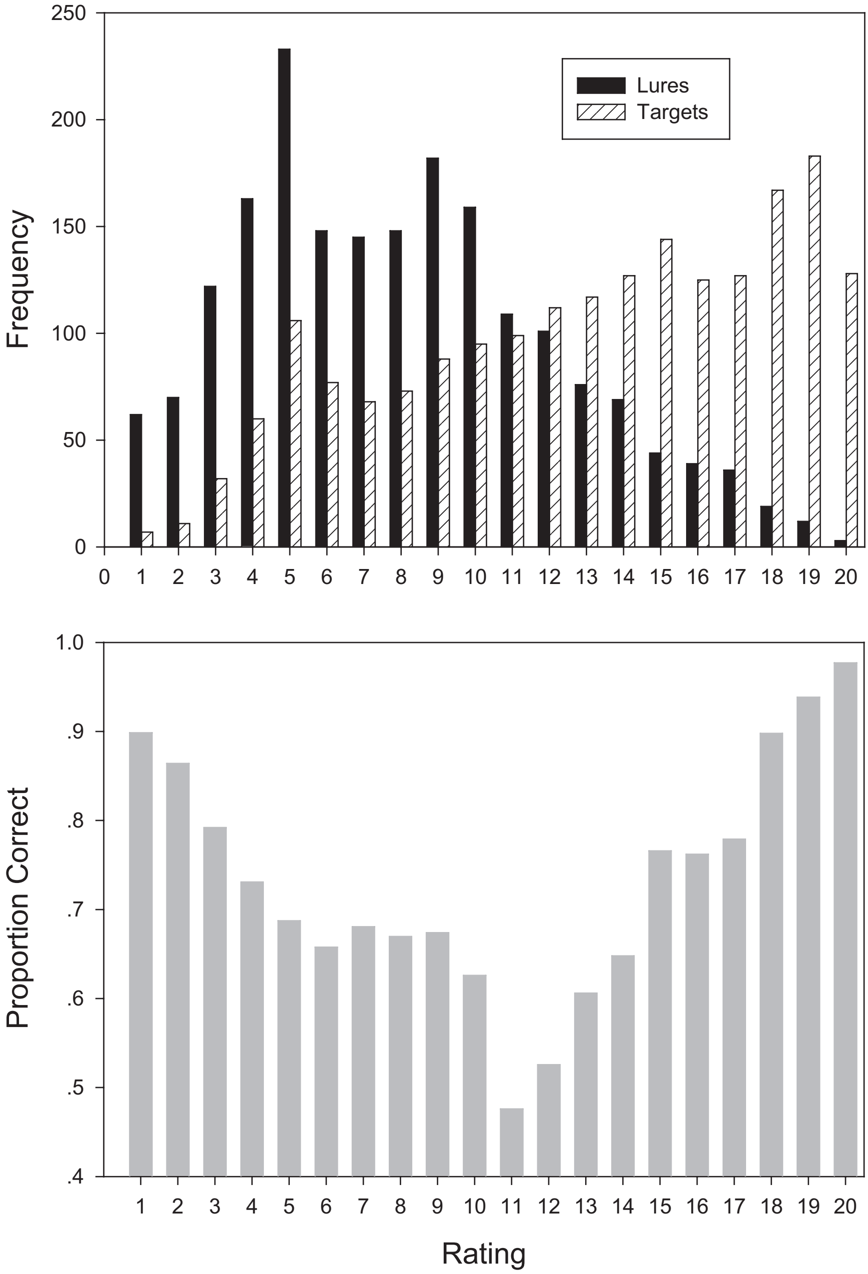

Figure 2. Upper panel: Frequency distribution showing the number of responses made to targets and lures pooled over subjects. Lower panel: Accuracy associated with each rating based on the pooled data in the upper panel.

lures was 8.04. The corresponding standard deviations—which are the main measures of interest—were 4.62 and 3.83, respectively. Table 1 also shows, for each subject, the ratio of the standard deviation of the lure ratings to the standard deviation of the target rating ($s_{\text{lure}}/s_{\text{target}}$). Excluding the outlier, the mean ratio is 0.83, which is significantly less than 1.0 [$t(12) = 3.52$]. With the outlier included, the mean ratio is 0.87, which is still significantly less than 1.0 [$t(13) = 2.42$].

We next conducted an ROC analysis on these data by counting the number of responses to targets and number of responses to lures that exceeded the following cutoffs on the rating scale: 17, 14, 11, 8, 5, and 1. That is, we treated the rating scale as if it were a 6-point confidence scale, with a rating of 17 to 20 being regarded as a high-confident *old* response, a rating of 14 to 16 as a medium-confident *old* response, and so on down to ratings of 1 to 4, which were treated as high-confident *new* responses. The confidence scale is assumed to provide only an ordinal scale of measurement. That is, the high-confident *old* criterion is assumed to be higher on the memory strength scale than the medium-confident *old* criterion, but the

Table 1
The Mean Rating Made by Each Subject to All Test Items ($m_{overall}$) in Experiment 1, As Well As the Means and Standard Deviations of the Ratings Made to Targets (m_{target} and s_{target}, Respectively) and Lures (m_{lure} and s_{lure}, Respectively) and the Ratios of the Lure and Target Standard Deviations Obtained Directly From the Ratings (s_{lure}/s_{target}) and From a Separate ROC Analysis ($\sigma_{lure}/\sigma_{target}$) of the Same Data

Subject	$m_{overall}$	m_{target}	m_{lure}	s_{target}	s_{lure}	s_{lure}/s_{target}	$\sigma_{lure}/\sigma_{target}$
1	12.09	15.34	8.83	5.55	6.19	1.12	0.91
2	9.83	13.09	6.57	6.35	4.30	0.68	0.76
3	10.10	13.11	7.08	4.87	3.75	0.77	0.64
4	11.87	14.68	9.05	3.78	2.40	0.63	0.56
5	10.32	12.28	8.35	4.34	2.81	0.65	0.55
6	10.71	14.45	6.97	5.47	5.61	1.03	0.87
7	8.89	10.47	7.31	5.04	3.65	0.72	0.76
8	8.93	11.90	5.95	5.35	3.60	0.67	0.75
9	10.29	10.73	9.85	2.31	2.39	1.03	1.06
10	10.89	12.89	8.89	4.26	3.71	0.87	0.83
11	14.17	17.61	10.73	3.51	4.54	1.30	0.73
12	11.79	15.80	7.78	4.59	4.85	1.06	0.74
13	9.55	10.47	8.62	4.43	3.61	0.81	0.88
14	11.37	13.51	9.23	3.77	2.96	0.79	1.01
Mean	10.77	12.98	8.04	4.62	3.83	0.83	0.79

Note—Except in the first column, Subject 11's scores were excluded from the mean values.

distance between those two criteria need not be the same as the distance between the medium-confident criterion and the low-confident criterion. Because the direct rating method and the ROC method entail quite different assumptions, they need not agree in their conclusions (as illustrated in detail later).

The ROC analysis was performed by fitting the Gaussian detection model to the ROC data of each individual subject using maximum likelihood estimation. One of the parameters of the model is the ratio of the standard deviation of the lure distribution divided by the standard deviation of the target distribution ($\sigma_{lure}/\sigma_{target}$). The estimated value of that ratio for each subject is shown in Table 1. The mean value was 0.79, which is typical and is quite close to the value obtained from direct ratings (0.83). Figure 3 shows the scatterplot of ratio measures derived from the two procedures for each subject. It is clear for the figure that the estimates are in good agreement ($r = .61, p < .05$).

The ROC analysis also yielded a discriminability measure for each subject, and the corresponding version of that measure was also computed directly from the ratings. The typical detection-based discriminability measure is d', which is the distance between the means of the target and lure distributions in standard deviation units. That is, $d' = (\mu_{target} - \mu_{lure})/\sigma$, where σ is the standard deviation of both the target and lure distributions. When an unequal-variance model applies, a related (and better) measure is d_a, which is the distance between the means relative to the root-mean square of the target and lure standard deviations:

$$d_a = \frac{\mu_{target} - \mu_{lure}}{\sqrt{\left(\sigma^2_{target} + \sigma^2_{lure}\right)/2}}.$$

For each subject, d_a was estimated from ROC analysis. A value analogous to d_a, denoted d_r, was then computed for each subject directly from the ratings according to the following formula:

$$d_r = \frac{m_{target} - m_{lure}}{\sqrt{\left(s^2_{target} + s^2_{lure}\right)/2}}.$$

Remarkably, the discriminability estimates obtained from ROC analysis and from the ratings were nearly identical ($r = .99$, mean $d_a = 1.15$, mean $d_r = 1.17$).

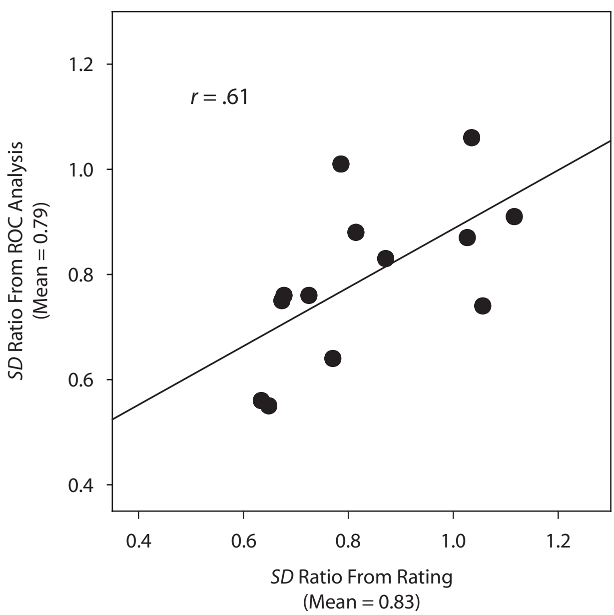

Figure 3. Scatterplot (and regression line) of lure-to-target standard deviation ratio estimates (ROC estimate vs. direct ratings estimate) from Experiment 1.

Discussion

ROC analyses of recognition memory almost invariably suggest that the memory strengths of the targets are more variable than the memory strengths of the lures. Using direct ratings of memory strength for targets and lures in which the means and standard deviations could be computed directly, we found that the results were in good agreement with ROC analysis. Both methods suggested that the standard deviation of the lure distribution is about .80 times that of the target distribution, on average, and the ratio estimates for individual subjects derived from the two methods correlated significantly.

The ratings method and the ROC method are not constrained to agree on this issue. To illustrate this, we conducted two simulations based on the signal detection models shown in the upper and lower panels of Figure 4. For both simulations, memory strengths for targets and lures were drawn from an equal-variance signal detection model with d' equal to 1.5. The two simulations differed only in how the 20-point rating scale was related to the underlying memory strength scale. In the first simulation, which is illustrated in the upper panel, the ratings were spread out on the weak end of the scale and compressed together on the strong end. As such, the difference in memory strength between ratings of 19 and 20 was small compared to the difference between ratings of 1 and 2. Thus, the rating scale did not have interval scale properties with respect to the psychological variable of interest (memory strength). The simulation involved drawing 150 memory strength values from the lure distribution and assigning a rating to each. Another 150 memory strength values were drawn from the target distribution, and ratings were assigned

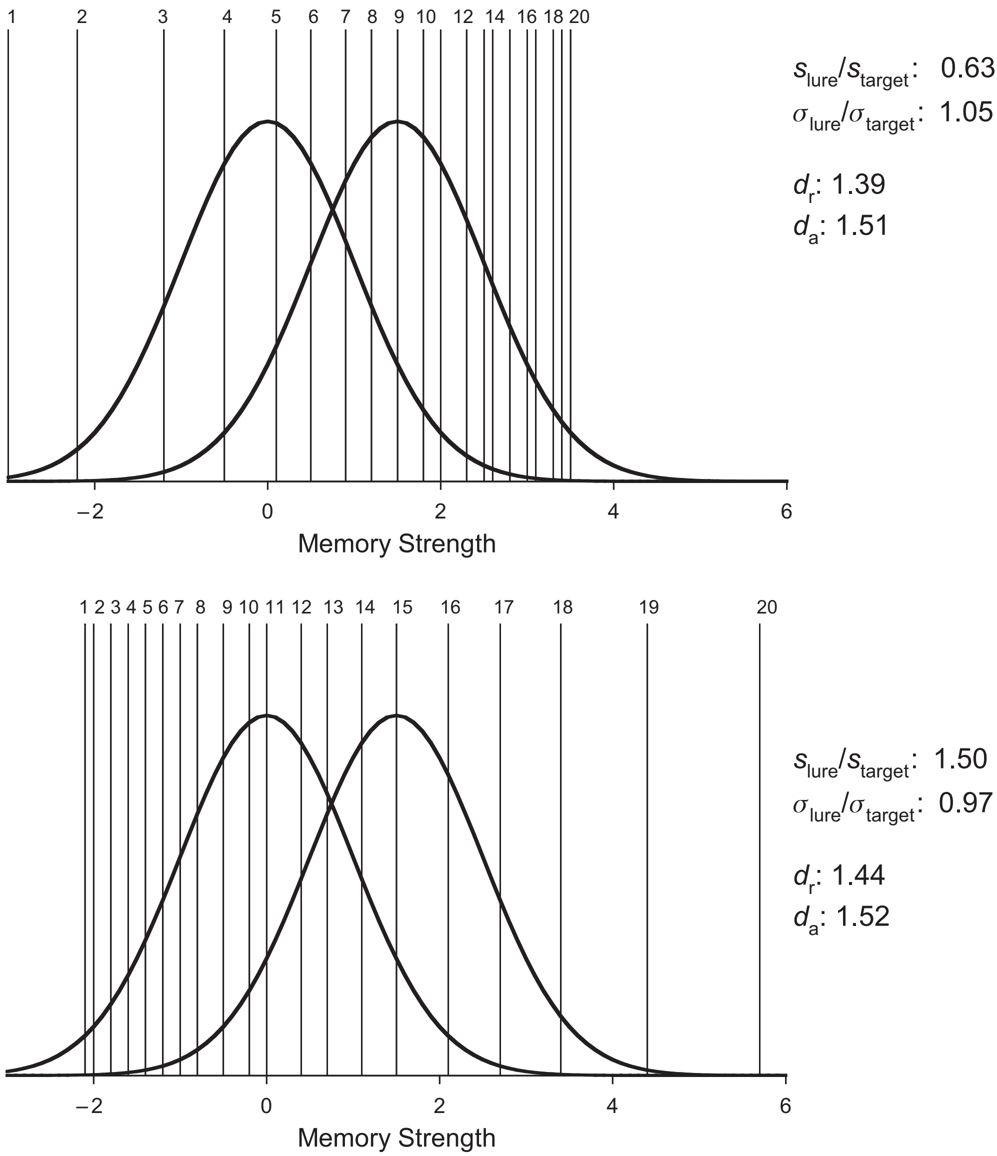

Figure 4. Hypothetical signal detection models illustrating two nonlinear relationships between a 20-point rating scale and the memory strength scale.

to them in the same way. The resulting rating data were then analyzed exactly as we analyzed the data presented above. That is, means and standard deviations were computed directly from the ratings and from ROC analysis. The ROC analysis yielded a standard deviation ratio close to the true value of 1 (namely, 1.05), which simply shows that ROC analysis is not dependent on the assumption of a linear measurement scale associated with the confidence ratings. By contrast, the direct-ratings method yielded an answer that was far off the mark (0.63 in this case) because it *is* dependent on the assumption of a linear measurement scale associated with the ratings (and that scale is intentionally nonlinear in this simulation).

The simulation was repeated using the equal-variance model shown in the lower panel of Figure 4. This time, the nonlinear relationship between scale ratings and memory strength was reversed, such that the difference in memory strength between a rating of 19 and a rating of 20 was very large in comparison with the difference in memory strength between a rating of 1 and a rating of 2. Once again, the ROC analysis returned a ratio estimate close to the true value of 1 (0.97), but the ratings returned an answer that was far off the mark (1.50), this time in the other direction. Wheras the slope estimate is very sensitive to the nature of the measurement scale, the estimates of d_r are much less affected.

EXPERIMENT 2

Although the results of Experiment 1 showed a correspondence between the ratio estimates derived from ROC analysis and from direct ratings of memory strength, the strength of the relationship may have been reduced by some rating anomalies that we attempted to eliminate in a second experiment. The ratings for 1 subject, for example, were clearly influenced by the old/new question that preceded the rating. This subject tended to avoid using the midrange of the scale and gave fairly high ratings to all items declared to be old and fairly low ratings to all items declared to be new. Thus, in Experiment 2, we eliminated the old/new question and asked for ratings only, this time using a 1–99 rating scale.

Method

Subjects. Sixteen undergraduates from University of California, San Diego, participated for lower-division psychology course credit.

Materials and Design. The words, list length, and duration of presentation were the same as those presented in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except that subjects did not make an initial old/new decision, they were informed that half of the words on the test were on the list presented and half were not, and they indicated the strength of their memory on a 1–99 scale.

Results

Scale biases were more apparent using the 1–99 scale. For example, subjects often supplied ratings at intervals of 5 on the scale, which means that, for them, this was effectively a 20-point scale, and there was a noticeable bias to choose the midpoint rating of 50 for both targets and lures. In addition, as in Experiment 1, the target distribution showed evidence of a ceiling effect, with 11.6% of the targets (and virtually none of the lures) receiving a rating of 99. Otherwise, the distribution and accuracy data were similar to the results of Experiment 1. The subjects were largely unbiased in the use of the scale, so that their ratings for all items averaged together (targets and lures) were close to the midpoint of the scale, with the mean value being 50.99. All of the scores were symmetrically distributed about 50 (ranging from 38.63 to 60.60), with no apparent outliers.

Table 2 shows the mean and standard deviations for the ratings made to the targets and to the lures for each subject. Across all 16 subjects, the mean rating for the

Table 2
The Mean Rating Made by Each Subject to All Test Items ($m_{overall}$) in Experiment 2, As Well As the Means and Standard Deviations of the Ratings Made to Targets (m_{target} and s_{target}, Respectively) and Lures (m_{lure} and s_{lure}, Respectively) and the Ratios of the Lure and Target Standard Deviations Obtained Directly From the Ratings (s_{lure}/s_{target}) and From a Separate ROC Analysis ($\sigma_{lure}/\sigma_{target}$) of the Same Data

Subject	$m_{overall}$	m_{target}	m_{lure}	s_{target}	s_{lure}	s_{lure}/s_{target}	$\sigma_{lure}/\sigma_{target}$
1	56.62	64.63	48.60	18.36	12.63	0.69	0.81
2	51.92	67.62	36.21	24.79	17.49	0.71	0.66
3	43.91	70.27	17.54	36.26	23.42	0.65	0.56
4	53.47	66.23	40.71	32.74	33.76	1.03	0.95
5	44.88	54.07	35.68	23.84	19.16	0.80	0.78
6	56.34	68.79	43.88	28.00	27.91	1.00	0.81
7	58.09	65.27	50.91	20.59	18.39	0.89	0.82
8	44.32	73.07	15.57	30.56	14.14	0.46	0.63
9	60.60	79.31	41.89	22.55	23.29	1.03	0.79
10	43.49	62.11	24.87	33.48	21.00	0.63	0.62
11	51.40	55.14	47.66	40.70	41.24	1.01	1.18
12	56.99	65.33	48.65	18.29	6.60	0.36	0.55
13	38.63	51.33	25.92	28.89	12.89	0.45	0.70
14	60.35	74.58	46.11	20.53	25.22	1.23	1.11
15	49.15	58.50	39.80	19.05	14.31	0.75	0.75
16	45.71	60.22	31.19	31.11	18.40	0.59	0.64
Mean	50.99	64.78	37.20	26.86	20.61	0.77	0.77

targets was 64.78 on the 99-point scale, and the mean rating for the lures was 37.20. Also shown for each subject is the ratio of the standard deviation of the lure ratings to the standard deviation of the target rating (i.e., $s_{\text{lure}}/s_{\text{target}}$). The mean ratio is 0.77, which is significantly less than 1.0 [$t(15) = 3.76$]. We next conducted an ROC analysis on these data by tabulating the number of responses to targets and number of responses to lures that exceeded the following cutoffs on the rating scale: 83, 67, 51, 33, 17, and 1. The estimated $\sigma_{\text{lure}}/\sigma_{\text{target}}$ ratio values for each subject are also shown in Table 2. The mean value of that ratio was also 0.77. Figure 5 shows the scatterplot of ratio measures derived from the two procedures, and the level of agreement is even higher than it was in Experiment 1 ($r = .83, p < .001$).

Finally, as in Experiment 1, the values of d_a estimated from the ROC analysis were remarkably similar to the d_r values estimated directly from the ratings ($r > .99$, mean $d_a = 1.12$, mean $d_r = 1.18$). It has been argued that, in the unequal-variance situation, d_a is the single best estimate of discriminability (Macmillan & Creelman, 2005), but it has never been widely used in the recognition literature because an ROC analysis was needed to obtain an estimate of it. It seems that a simpler way to obtain that estimate is to compute it directly from ratings of memory strength—ratings that are as easy to obtain as old/new decisions are.

GENERAL DISCUSSION

The two experiments reported here support a conclusion that is commonly drawn from ROC analysis—namely, that the memory strengths of the targets are more variable than the memory strengths of the lures. Using direct ratings of memory strength on a 1–20 scale or a 1–99 scale, we found that the standard deviation of the lure ratings was about 0.80 times the standard deviation of the target ratings. This is the predicted result given that the slope of the *z*-ROC is often approximately 0.80. Also, across individual subjects, ratio estimates derived from the direct-ratings method correlated highly with ratio estimates derived from ROC analysis. These two methods are not constrained to agree, and they rely on different assumptions. The ROC analysis relies on the assumption that the memory strength distributions are Gaussian in form. That assumption makes it possible to avoid the assumption that confidence ratings are made on a linear scale. The direct-ratings method, by contrast, assumes a linear scale, and so avoids having to make any assumption about the mathematical form of the distribution of memory strengths. Even so, the level of agreement between the two methods is remarkably high.

The close agreement between the model-based ROC analysis and the model-free ratings method supports not only an unequal-variance model, but also the idea that the memory strengths are distributed in such a way that fitting a specifically Gaussian model to the data yields accurate conclusions (even if the true underlying distributions are not strictly Gaussian). However, as indicated earlier, the subjects tended to choose the highest rating for about 10% of the targets, which might indicate that the target distribution has a long tail that extends well beyond the highest rating. If so, the estimated difference in variance between the targets and lures based on the ratings (but not the ROC analysis) would have been even greater. In that case, both methods would still support an unequal-variance model, but they would not agree on the degree of inequality.

The fact that quite a few targets but almost no lures received the highest rating in both experiments is consistent with the idea that only recollection gives rise to the highest memory strengths (in that recollection is likely to be associated with targets, not with lures). On the surface, this pattern might appear to suggest that recollection is an all-or-none phenomenon, but evidence weighing against this idea can be found in source memory studies showing that lower degrees of confidence are associated with lower degrees of recollective accuracy, not the absence of recollection (e.g., Slotnick & Dodson, 2005). Although we did not use a source memory procedure to test this idea here, it seems likely that varying degrees of recollective success were associated with different ratings of memory strength. As such, recollection is probably not represented solely in the highest rating, even though especially strong recollection may be responsible for the fact that only targets tend to receive that rating.

As noted by Wixted (2007), although it might seem that an unequal-variance model is inherently less plausible than the more aesthetically appealing equal-variance model, the opposite is actually true. The targets can be thought of as lures that have had memory strength added to them by virtue of their appearance on the study list. An equal-variance model would result if each item on the list

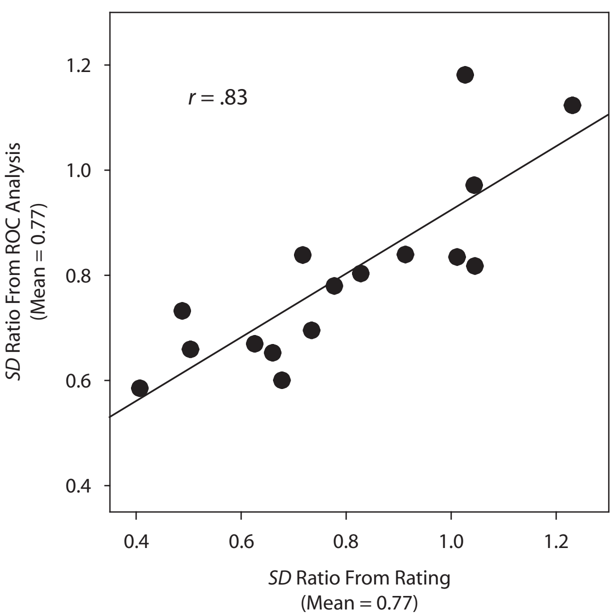

Figure 5. Scatterplot (and regression line) of lure-to-target standard deviation ratio estimates (ROC estimate vs. direct ratings estimate) from Experiment 2.

had the exact same amount of strength added during study. However, if the amount of strength that is added differs across items, as it must, then both strength and variability would be added, and an unequal-variance model would apply. It is, of course, possible to imagine forces that would work against the increased variance (e.g., if the amount of strength added during study is inversely proportional to baseline strength). However, because few would dispute the notion that varying degrees of strength are added at study, it is actually the equal-variance model that is, a priori, the less plausible account. The ratings data reported here suggest that the more plausible unequal-variance account, which has long been supported by ROC analysis, is substantiated by direct ratings of memory strength.

AUTHOR NOTE

We thank Norman Anderson for doubting the unequal-variance model and suggesting this experiment, and Janice Lee for her help in collecting the data. Correspondence concerning this article should be sent to J. T. Wixted, Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093 (e-mail: jwixted@ucsd.edu).

REFERENCES

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-505.

Egan, J. P. (1958). Recognition memory and the operating characteristic (Tech. Note AFCRC-TN-58-51). Bloomington: Indiana University, Hearing and Communication Laboratory.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. New York: Academic Press.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 500-513.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, **99**, 518-535.

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, **33**, 151-170.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, **114**, 152-176.

(Manuscript received June 22, 2006;
revision accepted for publication January 11, 2007.)