Journal of Experimental Psychology:
Learning, Memory, and Cognition
2007, Vol. 33, No. 1, 18–33

Copyright 2007 by the American Psychological Association
0278-7393/07/$12.00 DOI: 10.1037/0278-7393.33.1.18

The Mirror Effect and Mixture Signal Detection Theory

Lawrence T. DeCarlo
Columbia University

The *mirror effect* for word frequency refers to the finding that low-frequency words have higher hit rates and lower false alarm rates than high-frequency words. This result is typically interpreted in terms of conventional signal detection theory (SDT), in which case it indicates that the order of the underlying old item distributions mirrors the order of the new item distributions. However, when viewed in terms of a mixture version of SDT, the order of hits and false alarms does not necessarily imply the same order in the underlying distributions because of possible effects of mixing. A reversal in underlying distributions did not appear for fits of mixture SDT models to data from 4 experiments.

Keywords: mirror effect, signal detection theory, mixture models

The *mirror effect* (Glanzer & Adams, 1985, 1990) is a general phenomenon in recognition memory that has been studied with a number of variables, most often with word frequency. In that case, the mirror effect refers to the finding that low-frequency words have higher hit rates and lower false alarm rates than high-frequency words. This result is usually interpreted in terms of (unequal variance) signal detection theory (SDT), in which case it implies that the order of the underlying old item distributions mirrors the order of the new item distributions, just like the observed hit and false alarm rates.

Figure 1 illustrates the interpretation of the mirror effect in terms of conventional SDT; note that the relative spacing and relative variances of the normal distributions shown in the figure are consistent with those found below (with the exception that the distributions in Figure 1 are spaced further apart for visual clarity). The figure shows that the mirror effect consists of two fundamental aspects (Glanzer & Adams, 1985). The first is that the distribution for high-frequency new (HN) words is to the right of that for low-frequency new (LN) words. This suggests that HN words are more familiar than LN words, as expected because high-frequency words are more commonly encountered. A second aspect of the mirror effect is that the distribution for low-frequency old (LO) words is to the right of that for high-frequency old (HO) words. Thus, when one studies the words, the less familiar low-frequency words are strengthened more than the high-frequency words, with the result that the order of old item distributions mirrors that for new item distributions (i.e., the order is reversed); note that the SDT distance parameter d' is therefore larger for low-frequency words than for high-frequency words.

It should be recognized that in order for the mirror effect to hold, both aspects noted above must be present—that is, the HN word distribution must be to the right of the LN distribution and the LO distribution must be to the right of the HO distribution. Simply finding a larger value of d' for low-frequency words is not sufficient to demonstrate the mirror effect, as was noted by Glanzer and Adams (1990), who gave examples in which the mirror effect does not appear. In one example, d' was larger for low-frequency words, but the high- and low-frequency old item distributions had the same location; in a second example, d' was again larger, but the high- and low-frequency new item distributions had the same location.

The mirror effect has received considerable attention in memory research (e.g., Arndt & Reder, 2002; Benjamin, 2003; Dobbins & Kroll, 2005; Glanzer & Adams, 1985, 1990; Hintzman, 1988, 1994; Maddox & Estes, 1997; Malmberg, Holden, & Shiffrin, 2004; Malmberg & Nelson, 2003; McClelland & Chappell, 1998; Murdock, 1998, 2003; Park, Reder, & Dickison, 2005; Reder et al., 2000; Shiffrin & Steyvers, 1997; Sikström, 2001; Stretch & Wixted, 1998; Wixted, 1992), for several reasons. For one, the mirror effect is inconsistent with results found for recall, in that recall is typically better for high-frequency words than for low-frequency words (e.g., Gregg, 1976; Lockhart, Craik, & Jacoby, 1976; also see Overschelde, 2002). If high-frequency words are recalled better than low-frequency words, then why are they less well recognized? It has also been noted that the mirror effect is problematic for simple strength models of recognition memory (e.g., Glanzer & Adams, 1990; Hintzman, 1994; Murdock, 2003). As shown in Figure 1, something more than a simple strengthening (i.e., an increase in familiarity) apparently occurs when high- and low-frequency words are studied. The challenge is to explain why, upon study, the low-frequency words leapfrog over the high-frequency words.

Many researchers have suggested that the mirror effect reflects the influence of another process in recognition memory. Processes related to the level of processing or attention have often been appealed to. For example, Lockhart et al. (1976) and others (e.g., Hintzman, 1988; Rao & Proctor, 1984) have suggested that low-frequency words are processed at a deeper level than high-frequency words. In a similar vein, many researchers have suggested that low-frequency words receive more attention than high-frequency words (e.g., Glanzer & Adams, 1990; Maddox & Estes, 1997; Malmberg & Nelson, 2003; Mandler, 1980). The exact effects of the differential processing or attention, however, depend on the particular theory. For example, Glanzer and Adams (1990)

Correspondence concerning this article should be addressed to Lawrence T. DeCarlo, Department of Human Development, Teachers College, Columbia University, Box 118, 525 West 120th Street, New York, NY 10027. E-mail: decarlo@tc.edu

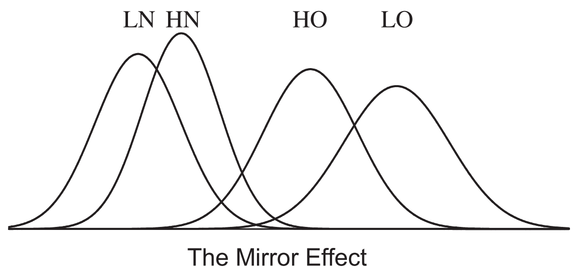

Figure 1. An illustration of the mirror effect with normal underlying distributions. LN = low-frequency new words; HN = high-frequency new words; HO = high-frequency old words; LO = low-frequency old words.

offered an attention-likelihood theory in which it is assumed that, among other things, the underlying distributions are binomial and participants use a likelihood-ratio decision rule. The complexity of this model (e.g., the use of a likelihood decision rule) has been criticized by several researchers (Hintzman, 1994; Murdock, 1998); it is shown here that neither of these assumptions (use of a likelihood ratio rule, binomial distributions) is needed to account for the mirror effect.

The present article shows that a simple generalization of SDT, introduced in recent articles, provides a new perspective on the mirror effect. The extended SDT model includes a parameter that allows for mixing over trials; for word recognition studies, it has been suggested that the mixing parameter can be interpreted as a measure of the level of processing or attention to the study words, and so the model offers the possibility of disentangling effects of attention from those of strength. The results found below are in fact consistent with earlier speculations about the possible influence of attention in recognition memory experiments.

Mixture SDT

The Mixture Model

Given the above ideas about the role of attention or level of processing in the mirror effect, a recent extension of SDT, namely a mixture extension (DeCarlo, 2000, 2002, 2003a), is clearly relevant. Mixture SDT is motivated by the view that there is an additional process involved in SDT that leads to mixing over trials. For example, in recognition memory research, one possibility is that some of the words were not attended to, or processed as deeply, during the study period. As a result, during the test trials, old words that are presented actually consist of two latent (unobserved) classes of words, such as attended and nonattended (or partially attended) old words. This in turn means that the distribution associated with old words in a conventional SDT analysis actually consists of a mixture of two distributions (because the analysis does not separate the attended from the nonattended words).

The upper panel of Figure 2 shows an example in which N is the underlying distribution for new words, O is the distribution for attended old words, and O′ is the distribution for partially attended old words; the O and O′ distributions represent words that were attended to at different levels (or processed at different levels) during the study period. The arrows above the distributions indicate that the O and O′ distributions are mixed over old word trials; that is, the old word distribution is located at O for a certain proportion of trials and at O′ for the remaining proportion of trials. Note that the normal distributions shown in the top panel of Figure 2 all have equal variance: In the mixture SDT approach, an increase in familiarity is represented by a simple shift in the location of the underlying distribution (and not as a simultaneous change in the variance; for notes on problems with the unequal variance approach, see DeCarlo, 2002). An attractive aspect of mixture SDT is that mixing can often be easily theoretically motivated, in that many processes (such as different levels of attention or processing) quite naturally lead to latent classes, which in turn leads to mixing over trials.

If the mixture SDT model is fit to (rating response) data, then one obtains estimates of the locations of the two unmixed distributions (O and O′) shown in the top panel of Figure 2; estimates of the mixing proportion and the locations of the response criteria are also obtained. On the other hand, if the unequal variance SDT model is fit to the data, then an estimate of the location (and relative variance) of a mixed old word distribution is obtained, which is shown in the lower panel of Figure 2 (note that the figure shows the density for an actual mixture of normal distributions). The mixed old word distribution results from the mixing over trials of the two old word distributions shown in the top panel of Figure 2.

Note that the mixed distribution in the lower panel of Figure 2 shows that mixing shifts the apparent location of the (attended) old word distribution (O in the top panel) to the left. The figure also shows that the mixed old word distribution has larger variance than the unmixed old word distributions shown in the top panel (it is also skewed). This is important because it shows that the mixture SDT model provides a theoretical account of the typical finding (from the unequal variance SDT view) that old word distributions have larger variance than new word distributions—according to mix-

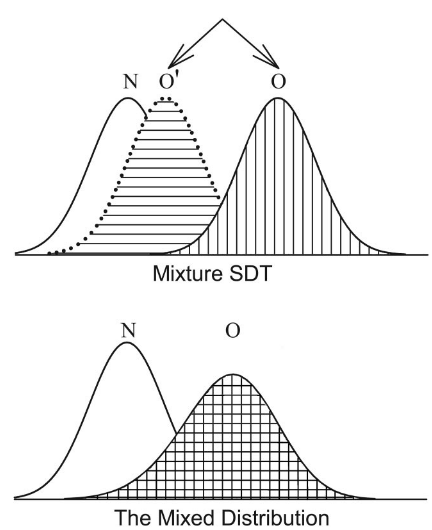

Figure 2. The top panel shows normal distributions associated with the mixture signal detection theory (SDT) model, with N indicating the new word distribution, O the attended old word distribution, O′ the partially attended distribution, and the arrows indicating which distributions are mixed. The bottom panel shows the mixed old word distribution that results from mixing in the top panel.

ture SDT, larger variance is found in conventional SDT because the old word distribution actually consists of a mixture of distributions (e.g., attended and nonattended words). A mixture SDT analysis in essence decomposes the larger variance (and possibly nonnormal) old word distribution of conventional SDT (lower panel) into two equal variance normal distributions (top panel).

Evidence of Validity

There are of course many possible reasons why one might find mixing in SDT. It has previously been shown, however, that results from several experiments support the view that the mixing parameter can be interpreted as a measure of attention or level of processing. For example, a mixture SDT analysis of published data of Ratcliff, McKoon, and Tindall (1994; DeCarlo, 2002) showed that the mixing parameter behaved as expected when variables that one would expect to affect attention were manipulated. The presentation time of the study words was varied in several experiments of Ratcliff et al. (1994), for example, and a mixture SDT analysis showed that the mixing parameter was systematically larger for longer presentation times. This result has a simple interpretation in terms of mixture SDT: The probability that a word is attended to is higher for longer presentation times, with the result that a greater proportion of study words are attended to, which in turn is reflected by a larger mixing parameter. Other experiments in which presentation time was manipulated have found similar results. For example, in source discrimination experiments (DeCarlo, 2003a), it was found that the mixing parameter again tended to be larger for longer presentation times.

It has also been shown that the mixing parameter varies with word frequency. For example, fits of mixture SDT models to data of Ratcliff et al. (1994) from a between-conditions manipulation of word frequency showed that the mixing parameter was consistently larger for low-frequency words than for high-frequency words (see DeCarlo, 2002). This is evidence that is consistent with the view that low-frequency words receive greater attention or are processed more deeply than high-frequency words, as many researchers have speculated. Thus, prior research has provided some evidence as to the validity of the interpretation of the mixing parameter as a measure of attention, although more work along these lines clearly needs to be done. The present experiments provide further evidence that the mixing parameter varies systematically with word frequency.

Another result of a mixture analysis of Ratcliff et al.'s (1994) word frequency data was that estimates of d' were larger for low-frequency words than for high-frequency words. This suggests that the mirror effect was found; however, strictly speaking, one cannot make this conclusion on the basis of a between-conditions manipulation (i.e., separate conditions with high- and low-frequency words, sometimes referred to as "pure" lists), because the mirror effect can be properly studied only by a within-condition manipulation (i.e., mixed lists)—that is, high- and low-frequency words should be presented in the same session.

There are actually several problems with a between-conditions manipulation. One is that although data from a high-frequency condition allow one to locate HO words relative to HN words, and data from a low-frequency condition allow one to locate LO words relative to LN words, one cannot locate HN words relative to LN words, or HO words relative to LO words, both of which are necessary in order to demonstrate the mirror effect. For the same reason, one cannot rule out the possibility that the response criteria differ across the conditions, as some have argued (e.g., Dobbins & Kroll, 2005; Hirshman, 1995; Stretch & Wixted, 1998). Another problem is that the underlying distributions are not necessarily scaled the same across the two conditions. That is, if one estimates the detection parameter d' or plots separate receiver operating characteristic (ROC) curves for each condition, as is often done, then one cannot compare the values of d' across the two conditions without further assuming that the variances of the reference distributions (LN and HN) are the same, which does not appear to be the case (as shown below). These problems do not arise if both high- and low-frequency words are presented in the same condition (i.e., using mixed lists), because all of the distributions can then be located and scaled to a common reference distribution (and there will be three ROC curves on an ROC plot, and not separate ROC plots for high- and low-frequency words, as are often presented). Glanzer and Adams (1990) made similar points in a discussion of between-subjects comparisons.

Although mixture SDT is clearly relevant to research on the mirror effect, I know of no studies where mixture SDT models have been fit to the relevant data: studies where both high- and low-frequency words were presented in the same session and a rating response was used (note that neither the variances in conventional SDT nor the mixing proportions in mixture SDT can be estimated from single-session binary response data). Because of the lack of available published data (i.e., frequencies for rating response experiments), an experiment was conducted to collect data on the mirror effect. In addition, rating response data from a published study of the mirror effect (Arndt & Reder, 2002) were obtained. The data from both experiments are examined in terms of mixture SDT models. Although the focus here is on mixture SDT models, results for the unequal variance SDT model are also presented to allow comparisons with earlier research; in addition, full results for a fit of the unequal variance SDT model (i.e., fit statistics, standard errors) have never been reported, to my knowledge, in spite of extensive research on the mirror effect. The model-based approach to the mirror effect offered here provides a detailed and unified approach to the data and offers a useful framework for accumulating knowledge across studies.

Mixture SDT Models for the Mirror Effect

The unequal variance SDT model is well known and so is not presented in detail; for a presentation of the model and statistical analysis, see DeCarlo (2003b). Mixture SDT models as applied to the mirror effect have not been presented before, and so the models are shown here.

The Mixture 1 SDT Model

The first mixture SDT model considered is the same as the one presented in DeCarlo (2002; Equation 1) and illustrated in Figure 2, except that it is applied to the mirror effect. Thus, there are two attended old distributions (for high- and low-frequency old words) and two partially attended old distributions (again for high- and low-frequency old words). It follows that a normal mixture SDT model for the mirror effect is

$$p(Y \le k|\boldsymbol{X},W,V) = \Phi(c_k - d_{\mathrm{LN}}X_1 - d_{\mathrm{HO}'}X_2 - \Delta d_{\mathrm{HO}}X_2W - d_{\mathrm{LO}'}X_3 - \Delta d_{\mathrm{LO}}X_3V), \quad (1)$$

where Y is the response variable with values k from 1 to K; X_1, X_2, and X_3 are dummy-coded variables that indicate LN, HO, and LO words, respectively (HN is used as the reference distribution); W and V are latent dummy-coded variables (indicating partial-attention/attention for high- and low-frequency old words, respectively); and Φ is the cumulative distribution function for the normal distribution. Note that the subscripts on d indicate the distribution it represents the distance for (e.g., $d_{\mathrm{HO}'}$ is the distance of the partially attended HO′ distribution from the reference HN distribution), whereas the deltas indicate that the parameter represents the change in distance from the corresponding partially attended distribution (e.g., Δd_{HO} indicates the distance of the attended HO distribution from the partially attended HO′ distribution). Thus, one can obtain d for the attended distribution by simply adding (e.g., $d_{\mathrm{HO}} = d_{\mathrm{HO}'} + \Delta d_{\mathrm{HO}}$); the standard errors can be obtained by noting that $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$, which was used for the results reported below.

The model is completed by embedding the SDT model into a restricted latent class model (see Dayton, 1998; DeCarlo, 2002),

$$p(Y \le k|\boldsymbol{X}) = \textstyle\sum_W \sum_V p(W)p(V)p(Y \le k|\boldsymbol{X},W,V), \quad (2)$$

where $p(W)$ and $p(V)$ are mixing proportions, written more compactly below as $\lambda_{\mathrm{HO}} = p(W)$ and $\lambda_{\mathrm{LO}} = p(V)$ for the high- and low-frequency words, respectively, and $p(Y \le k \mid \boldsymbol{X}, W, V)$ is as given in Equation 1. Equation 2 shows that the cumulative response probabilities for the observed variables $\boldsymbol{X}$ (i.e., the word types) are the sum of the product of the marginal and conditional probabilities over the latent classes of W and V. It is also assumed that the probabilities of attending to high- and low-frequency words are independent (this can be relaxed but is not theoretically motivated at this point and also did not improve fit in any case). Further details on mixture SDT models can be found in DeCarlo (2002).

The Mixture 2 SDT Model

At the outset, it was expected that the Mixture 1 model given above would not be fully satisfactory because it cannot account for the larger variance of the LN distribution relative to the HN distribution, shown in Figure 1 and found in Table 3 later (which shows that it is still useful to start with a conventional SDT analysis). The larger variance for the LN distribution suggests the possibility of mixing for LN words, as do other considerations. For example, several participants said things such as "I knew the words I didn't know the meaning of were probably old words," which is an incorrect inference because the test trials included low-frequency new words. Nevertheless, this comment suggests that some of the LN words were treated as old (e.g., falsely recognized), which would result in mixing for LN words—that is, some of the low-frequency new words were treated as old (denoted below as LN* words), possibly because participants treated words whose meaning they did not know (or words that were simply unusual) as being old, whereas others were (correctly) treated as new (LN words).[1]

The idea of mixing for LN words has also in essence previously been considered in the recognition memory literature. For example, to explain a high hit rate for rare words, Wixted (1992) noted that participants often do not know the meaning of rare words and therefore might encode them differently, with the result that "lures may be falsely recognized because of their perceptual or semantic similarity to encoded targets" (p. 689). This is essentially the type of account being offered here, with the difference that it is assumed that only a portion (and not all) of the LN words were "falsely recognized." This possibility is examined in more detail below in an experiment in which similar new words were purposely constructed (Arndt & Reder, 2002, Experiment 1).

To allow for mixing of LN words, Equation 1 is extended by adding one additional term, which is an interaction term (i.e., X_1U),

$$p(Y \le k|\boldsymbol{X},W,V,U) = \Phi(c_k - d_{\mathrm{LN}}X_1 - \Delta d_{\mathrm{LN}^*}X_1U - d_{\mathrm{HO}'}X_2 - \Delta d_{\mathrm{HO}}X_2W - d_{\mathrm{LO}'}X_3 - \Delta d_{\mathrm{LO}}X_3V), \quad (3)$$

where LN* indicates the distribution for "similar" (i.e., falsely recognized) low-frequency new words and U is a latent dummy categorical variable, with 1 indicating that the word is taken as being similar to an old word and 0 indicating that it is not (note that the coefficient Δd_{LN^*} indicates the change in location of the LN* distribution from the LN distribution). Equation 2 is also extended by including the mixing proportion $p(U)$,

$$p(Y \le k|\boldsymbol{X}) = \textstyle\sum_W \sum_V \sum_U p(W)p(V)p(U)p(Y \le k|\boldsymbol{X},W,V,U). \quad (4)$$

The mixing parameter $p(U)$ is denoted below as λ_{LN^*} and can be interpreted as indicating the proportion of low-frequency new words that are treated like (confused with) low-frequency old words (i.e., λ_{LN^*} is the proportion of falsely recognized low-frequency new words), possibly because their meaning was not known or because of surface similarity to low-frequency old words. Equations 3 and 4 generalize the basic mixture model of Equation 1 by allowing for an additional mixing process for LN words, namely that some of them are treated like old words. Evidence in favor of this interpretation would be the finding that the distribution of falsely recognized low-frequency new words (LN* words) is (far) to the right of the non-falsely-recognized low-frequency new words (LN words).

Data for the Mirror Effect

Method

Experiments 1a and 1b

Participants. Analysis of data supplied by Arndt and Reder (2002) suggested that a large sample size would likely be needed to obtain adequate parameter estimation (discussed below). For Experiment 1a, 72 graduate students from courses at Teachers College, whose native language was English, served as participants; participation was voluntary. Each student took part in one session, which was about 20 min in duration.

[1] One can also allow for mixing of HN words (or even both LN and HN words); however, the results did not suggest a need for this.

Table 1
Response Frequencies for Experiments 1a and 1b

Word type	1	2	3	4	5	6	Total
Experiment 1a (72 native speakers)							
Low-frequency new	1,457	1,436	991	567	398	191	5,040
High-frequency new	721	1,418	1,460	860	422	159	5,040
Low-frequency old	233	457	568	729	995	2,058	5,040
High-frequency old	211	689	1,016	1,028	894	1,202	5,040
Experiment 1b (42 nonnative speakers)							
Low-frequency new	899	755	504	344	236	202	2,940
High-frequency new	740	791	616	390	230	173	2,940
Low-frequency old	215	305	309	338	478	1295	2,940
High-frequency old	317	449	448	374	405	947	2,940

Volunteers who were nonnative English speakers were also available. Rather than simply excluding these potential participants (as is often done), an effort was made to obtain as many as possible because it is of interest to see whether the mirror effect appears for nonnative speakers and to compare the results with those found for native speakers. For Experiment 1b, data for a total of 42 nonnative English speakers were collected.

Materials. The 280 words were selected from the MRC psycholinguistic database (http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm) with the criteria that the words were five to seven letters in length; 140 words (the high-frequency words) had a frequency count (Kučera & Francis, 1967) of greater than 100 per million (the range was 101–1,815, with a median of 202), and 140 words (the low-frequency words) had a frequency count of 1–9 per million (with a median of 3).

Design and procedure. The experiments were run on personal computers using E-prime (Version 1.1; Schneider, Eschman, & Zuccolotto, 2002). Participants were first given a short practice session, with five old words and five new words. In both the practice session and the test, participants rated their confidence that the word was old or new using a 1 to 6 response scale, with 1 = *sure new*, 2 = *fairly sure new*, 3 = *slightly sure new*, 4 = *slightly sure old*, 5 = *fairly sure old*, and 6 = *sure old*.

The study phase of the experiment consisted of two blocks of study–test during which a total of 280 test words were presented. In each block, 70 words were first presented for study in the center of the computer screen for 1 s; 35 of the words were high-frequency words, and 35 were low-frequency words. For the test in each block, 140 words were presented, of which 70 were the high- and low-frequency words that had been presented during the study period, 35 were high-frequency new words, and 35 were low-frequency new words. Participants entered their responses by using the numbered keys located at the top of the computer keyboard; participants were told that they should try to use each response category at least once.

Arndt and Reder (2002), Experiment 2

Also analyzed are mirror effect data from another experiment that used a 1 to 6 rating scale together with high- and low-frequency words (Arndt & Reder, 2002, Experiment 2).[2] The study consisted of 20 students, all native English-language speakers, who were each presented with a study list of 180 words (90 low-frequency and 90 high-frequency) presented for 2 s each and a test list of 360 words (180 old and 180 new). Low-frequency words had a frequency count of less than 4 per million; high-frequency words had a count of greater than 24 per million. Further details can be found in Arndt and Reder (2002).

Results and Discussion

Table 1 presents the frequencies for Experiments 1a and 1b pooled over subjects. The table shows that the basic data for a mirror effect study in this case consist of a 4 × 6 table, with 4 rows for the 4 word types and 6 columns for the 1–6 responses. The table also shows that the row totals are fixed by design. The unequal variance normal SDT model was fit by creating three zero–one dummy variables to indicate the four word types, with the high-frequency new words used as the reference, and then fitting a probit model using SPSS, as shown in Example 3 of DeCarlo (2003b). The mixture normal SDT models were fit using LEM (Vermunt, 1997; this software is freely available on the Web, and a link appears on my Web site: http://www.columbia.edu/~ld208); the models were also fit using aML (Lillard & Panis, 2003; also freely available) and Mplus Version 4.1 (Muthén & Muthén, 2006).

Unequal Variance SDT Model

Goodness of fit. The model being fit is the standard unequal variance SDT model with normal underlying distributions; see DeCarlo (2003b) for details. Table 2 presents, for Experiments 1a and 1b and Arndt and Reder's (2002) Experiment 2, absolute (a likelihood ratio test [LR]) and relative (Akaike information criterion [AIC] and Bayesian information criterion [BIC]) goodness-of-fit statistics for all of the models considered in this article. AIC and BIC are discussed in the section below on mixture models; the results for fits of a dual-process model are also shown and are discussed below. The LR goodness-of-fit statistics show that for all three studies, the unequal variance SDT model is rejected with respect to absolute fit.[3] Thus, the goodness-of-fit statistics suggest lack of fit for the unequal variance SDT model. Although the unequal variance SDT model has been commonly used in studies

[2] I thank Jason Arndt and Lynne Reder for supplying their data.

[3] Other links were also examined (see DeCarlo, 1998), but none gave smaller fit statistics.

Table 2
Goodness-of-Fit Statistics and Information Criteria for Experiments 1a, 1b, and Arndt and Reder's (2002) Experiment 2

Model	LR	*df*	*p*	AIC	BIC
Experiment 1a (N = 20,160)					
Unequal variance SDT	59.97	9	<.001	65,512	65,599
Mixture SDT (Mix 1)	207.69	8	<.001	65,661	65,756
Mixture SDT (Mix 2)	10.41	6	.108	65,468	65,579
Dual-process model	344.27	10	<.001	65,794	65,873
Experiment 1b (N = 11,760)					
Unequal variance SDT	22.05	9	.009	38,826	38,907
Mixture SDT (Mix 1)	32.27	8	<.001	38,838	38,927
Mixture SDT (Mix 2)	7.16	6	.306	38,817	38,920
Dual-process model	81.06	10	<.001	38,883	38,957
Arndt and Reder, Experiment 2 (N = 7,200)					
Unequal variance SDT	76.86	9	<.001	23,619	23,695
Mixture SDT (Mix 1)	23.77	8	.003	23,568	23,651
Mixture SDT (Mix 2)	11.04	6	.087	23,560	23,656
Dual-process Model	38.82	10	<.001	23,579	23,648

Note. LR = likelihood ratio goodness-of-fit test; AIC = Akaike information criterion; BIC = Bayesian information criterion; SDT = signal detection theory; Mix 1 = Equation 1; Mix 2 = Equation 3.

of the mirror effect, a lack of fit has not been noted up to this point because (LR or chi-square) fit statistics have previously not been reported.

As shown in Table 2, the sample sizes for the pooled data are rather large, and so one should keep in mind that the fit statistics likely have high power. To help ensure that one is not rejecting a model owing to a large sample size yet trivial deviations, it is useful to supplement the fit statistics by presenting an ROC plot ofthe data and fitted ROC curves; this allows one to visually assess deviations from the model. Figure 3 shows the data and fitted curves for Experiments 1a and 1b and Arndt and Reder's (2002) Experiment 2. The left panels show the data and fitted unequal variance SDT lines on inverse normal (i.e., z transformed) coordinates (i.e., the figures show z-ROC curves). Note that the LN curve is below the reference diagonal because the HN distribution is used as the reference distribution, and so the curve simply shows that the LN distribution is to the left of the HN distribution (any distribution can be used as the reference; however, for one of the mixture models considered here, the HN must be used as the reference, and so that is done for all of the models).

Figure 3 shows that for all three experiments, the deviations of the data from the fitted unequal variance z-ROC curves are generally small. One problem, however, is that the deviations do not appear to be random but tend to show a systematic pattern, in that the plots suggest nonlinearity. This is most evident for Arndt and Reder's (2002) data, shown in the bottom panels. For example, the black and gray circles in the left bottom panel show the LO and HO data, and it is apparent that the leftmost and rightmost circles are above the fitted line, whereas those in between are below the fitted line. This suggests some upward curvature; the same pattern appears for the LN data. The curvature is more evident for the fitted mixture z-ROC curves, shown in the bottom right panel and discussed below. Nonlinearity also appears in the data of Experiments 1a and 1b (top and middle panels). For example, the gray circles for HO data show some upward curvature in both Experiments 1a and 1b, and the black circles for LO data show curves with a dip in both experiments, which is a characteristic of mixture z-ROC curves (see DeCarlo, 2002, Figure 3; this appears again in Figure 5 later). Together, these results suggest that the goodness-of-fit statistics shown in Table 2 are large and significant because the unequal variance SDT model fails to describe nonlinearity that appears in z-ROC plots.

Parameter estimates. Table 3 presents parameter estimates and standard errors for fits of the unequal variance SDT model to the data of Experiments 1a and 1b and Arndt and Reder's (2002) Experiment 2. The table shows that the mirror effect is clearly

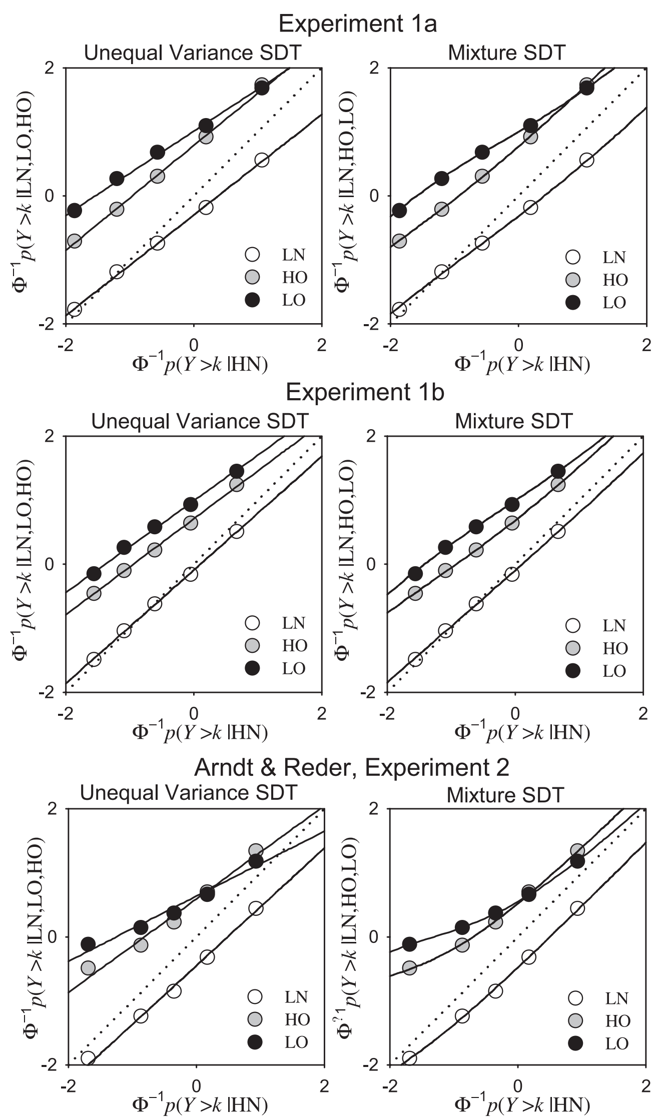

Figure 3. Data and fitted z-transformed receiver operating characteristic (z-ROC) curves for Experiment 1a, Experiment 1b, and Arndt and Reder's (2002) Experiment 2. SDT = signal detection theory model; LN = low-frequency new words; HN = high-frequency new words; HO = high-frequency old words; LO = low-frequency old words.

Table 3
Parameter Estimates for the Unequal Variance Signal Detection Model for Experiments 1a, 1b, and Arndt and Reder's (2002) Experiment 2

Parameter	LN	HN	HO	LO
		Experiment 1a		
d'	−0.38 (0.03)	0	0.96 (0.03)	1.53 (0.03)
log σ_s	0.24 (0.02)	0	0.20 (0.02)	0.42 (0.02)
		Experiment 1b		
d'	−0.10 (0.03)	0	0.94 (0.04)	1.38 (0.04)
log σ_s	0.12 (0.03)	0	0.29 (0.03)	0.33 (0.03)
		Arndt & Reder, Experiment 2		
d'	−0.49 (0.04)	0	0.81 (0.04)	1.25 (0.06)
log σ_s	0.09 (0.03)	0	0.31 (0.03)	0.68 (0.04)

Note. Standard errors are in parentheses. LN = low-frequency new words; HN = high-frequency new words; HO = high-frequency old words; LO = low-frequency old words.

evident, in that for all three experiments, the HN distribution is to the right of the LN distribution (as shown by the negative sign for LN) and the LO distribution is to the right of the HO distribution (and the standard errors are small), which are the two basic aspects of the mirror effect discussed above. The table also shows a pattern for the variances, with the variance for the HN distribution being the smallest, the variance for LO being the largest, and the variances of LN and HO in between. The pattern in variances is similar to that discussed by Glanzer and Adams (1990), in that it follows that the ratio of slopes (obtained by taking the inverse of the exponentiated value of log σ_S and then taking the appropriate ratios) is smallest for the LO/HN distributions and largest for the HO/LN distributions, as found by Glanzer and Adams, with the LO/LN and HO/HN ratios in between (for Experiments 1a and 1b, the order is HO/HN followed by LO/LN, whereas for Arndt and Reader's experiment, the order is LO/LN followed by HO/HN, as found by Glanzer & Adams, 1990). Thus, Table 3 suggests that in terms of the unequal variance SDT model, there is a consistent pattern in the locations of the distributions (i.e., the mirror effect) and also to some extent a pattern in the variances. As shown below, the mixture SDT model accounts for both of these results.

Mixture SDT Models

Goodness of fit. Table 2 shows absolute and relative fit statistics for the two mixture models described above. The LR statistics for the basic mixture model of Equation 1 are significant for all three experiments, indicating a lack of fit, as found above for the unequal variance SDT model. In contrast, the LR statistics for the mixture model of Equation 3 are considerably smaller and are not significant for Experiment 1a (in spite of the large sample size), Experiment 1b, or Arndt and Reder's (2002) Experiment 2, thus indicating that the Mixture 2 model adequately describes the data. For all three experiments, the information criterion AIC favors (i.e., is smallest for) the Mixture 2 model over the other models. For Experiment 1a, the information criterion BIC is smallest for the Mixture 2 model, but it is smallest for the unequal variance model for Experiment 1b and for the dual-process model for Arndt and Reder's experiment; this reflects the well-known result that BIC tends to favor less complex models than AIC (e.g., see Dayton, 1998) because it has a larger penalty for additional parameters.

The right panels of Figure 3 show fitted z-ROC curves for the Mixture 2 model. The figure shows that for all three experiments, the Mixture 2 SDT model accurately describes the deviations from linearity noted above (note that the Mixture 1 model generally fails to fit because the LN data show nonlinearity and a slope less than unity, whereas the Mixture 1 model predicts a linear curve with unit slope for LN words, as noted above).

Parameter estimates. Table 4 shows parameter estimates for the Mixture 2 model fit to the data of all three experiments using maximum likelihood estimation.[4] A simple and important result shown in Table 4 is that the estimate of the mixing parameter for low-frequency words (λ_{LO}) is larger in value than that for high-frequency words (λ_{HO}) for Experiment 1a (0.75 for λ_{LO} vs. 0.31 for λ_{HO}), Experiment 1b (0.74 for λ_{LO} vs. 0.43 for λ_{HO}), and Experiment 2 of Arndt and Reder (0.50 for λ_{LO} vs. 0.26 for λ_{HO}). This result can be interpreted as showing that a greater proportion of low-frequency old words were attended to than high-frequency old words. Thus, the present results join earlier results that suggest that low-frequency words receive greater attention than high-frequency words.

[4] Fits of the Mixture 1 model showed a similar pattern of results and so are not presented (as shown in Table 2, the fit of the model is also poor); that is, the basic four distributions were still ordered as LN, HN, LO, and HO. The main difference for fits of the Mixture 1 model was that the HO distribution tended to be further to the right of the LO distribution as compared with the Mixture 2 model; however, the standard errors were also large.

Table 4
Parameter Estimates for the Mixture 2 Signal Detection Model for Experiments 1a, 1b, and Arndt and Reder's (2002) Experiment 2

Model	LN	LO′	HN	HO′	LN*	LO	HO
				Experiment 1a			
Mix 2	−0.82 (0.07)	−0.16 (0.09)	0	0.49 (0.07)	0.74 (0.12)	1.98 (0.06)	2.10 (0.23)
		λ_{LO} = 0.75 (0.02)		λ_{HO} = 0.31 (0.06)		λ_{LN*} = 0.30 (0.04)	
				Experiment 1b			
Mix 2	−0.45 (0.22)	−0.04 (0.14)	0	0.22 (0.10)	0.60 (0.35)	1.80 (0.10)	1.93 (0.20)
		λ_{LO} = 0.74 (0.04)		λ_{HO} = 0.43 (0.07)		λ_{LN*} = 0.34 (0.23)	
				Arndt and Reder (2002), Experiment 2			
Mix 2	−0.54 (0.05)	−0.20 (0.07)	0	0.25 (0.04)	1.58 (0.69)	2.85 (0.16)	11.39 (—)
		λ_{LO} = 0.50 (0.02)		λ_{HO} = 0.26 (0.01)		λ_{LN*} = 0.04 (0.03)	

Note. The first row for each model shows estimates of *d* with standard errors in parentheses. The second row for each model shows the mixing parameters (λ_{LO} = low-frequency old; λ_{HO} = high-frequency old; λ_{LN*} = low-frequency new). LN = low-frequency new words; LO′ = unattended low-frequency old words; HN = high-frequency new words; HO′ = unattended high-frequency old words; LN* = similar low-frequency new words; LO = low-frequency old words; HO = high-frequency old words; Mix 2 = Equation 3.

With respect to the distance measures, it is interesting to note that the order of the seven distributions is exactly the same across all three experiments. For all three experiments, the new word distributions are ordered from left to right as LN and HN, as shown by the negative estimates of d_{LN} in Table 4, which is one aspect of the mirror effect. Also in all three experiments, the LO′ distribution is to the right of the LN distribution and the HO′ distribution is to the right of the HN distribution, which is consistent with the view that they are partially attended distributions (unattended distributions would simply have the same locations as the LN and HN distributions; the parameter estimates and standard errors in Table 4 show that this is generally not the case). It is also interesting to note that the LO′ distribution is close to (or to the left of) the HN distribution in all three studies; this suggests that partially attended low-frequency old words are about as familiar as high-frequency new words. Finally, there are no apparent differences between native and nonnative English speakers (it is interesting to note that for the nonnative speakers, the parameter estimates tend to be closer to zero, but the standard errors are relatively large).

An important result shown in Table 4 is that the ordering of the attended old word distributions is not HO and LO, as found for fits of the unequal variance SDT model, but rather, in all three experiments, the point estimates of *d* for the mixture model order the distributions as LO and HO. The differences in *d*s, however, are small (i.e., <0.20) for Experiments 1a and 1b (1.98 for LO vs. 2.10 for HO for Experiment 1a; 1.80 for LO vs. 1.93 for HO for Experiment 1b), and taken together with the standard errors, there is no evidence of any difference in the locations of LO and HO. A likelihood ratio test of the equality of *d* for HO and LO can be performed by fitting a restricted model where $d_{HO} = d_{LO}$, which in this case gives an LR statistic of 0.39 ($df = 1$, $p = .53$) for Experiment 1a and an LR of 0.67 ($df = 1$, $p = .41$) for Experiment 1b, and so the restriction is not rejected in either experiment. Thus, the results for Experiments 1a and 1b show that the mirror effect, although found for the observed response proportions, does not appear for the underlying distributions when mixture SDT models are fit. Instead, the results suggest that after study, low- and high-frequency words have the same familiarity.

Note that for Arndt and Reder's (2002) study, the standard error for the estimate of d_{HO} is indeterminate. This occurred because the fitted model was empirically poorly identified; that is, the ratio of largest to smallest eigenvalues of the information matrix (the information matrix is used to obtain the standard errors) was large, and so the standard errors for some parameters (in this case d_{HO}) were large or indeterminate. This generally occurs in SDT when the detection parameter is large; for example, in a simple 2 × 2 table, as *d* gets large the frequencies concentrate along the diagonal, and so the off-diagonal frequencies get smaller (i.e., the table gets sparser) and the standard errors get larger. In mixture SDT applied to the mirror effect, problems arise for HO in this case because the values of *d* appear to be relatively large (i.e., greater than 2) and because the mixing parameter is small; that is, the old word distribution is located at HO only 26% of the time, which leads to weak data (and large standard errors) with respect to determining the location of HO. Given the problem with identification for the data of Arndt and Reder (2002), a conservative conclusion is that it is not clear for Arndt and Reder's data whether HO and LO have the same location or HO is to the right of LO; in any case, neither outcome is consistent with a reversal in the order of high- and low-frequency distributions across new and old words.

A new result shown in Table 4 has to do with the unmixed low-frequency new word distributions. In particular, the parameter estimates in Table 4 show that for all three experiments, the unmixed LN* distribution is far to the right of the LN distribution (1.56 for Experiment 1a, 1.05 for Experiment 1b, and 2.12 for Arndt and Reder's 2002 experiment). This is an important result that is consistent with the interpretation that the LN* distribution represents low-frequency new words that are treated like LO words; a failure to find

Table 5
Mean Parameter Estimates for Individual Data for the Mixture 2 Signal Detection Model

Experiment	LN	LO′	HN	HO′	LN*	LO	HO
			Experiment 1a (n = 27)				
1a	−1.28	−0.46	0	0.25	1.41	2.81	2.61
		λ_{LO} = 0.73		λ_{HO} = 0.46		λ_{LN*} = 0.32	
			Experiment 1b (n = 12)				
1b	−1.19	−0.35	0	0.19	1.17	2.86	2.97
		λ_{LO} = 0.68		λ_{HO} = 0.43		λ_{LN*} = 0.40	

Note. The first row for each experiment shows estimates of d. The second row for each experiment shows the mixing parameters (λ_{LO} = low-frequency old; λ_{HO} = high-frequency old; λ_{LN*} = low-frequency new). LN = low-frequency new words; LO′ = unattended low-frequency old words; HN = high-frequency new words; HO′ = unattended high-frequency old words; LN* = similar low-frequency new words; LO = low-frequency old words; HO = high-frequency old words.

this result would cast doubt on the “false recognition” interpretation offered here. The mixing proportions in Table 4 show that the proportion of LN* words was about 30% in Experiment 1a, 34% in Experiment 1b, and 4% in Arndt and Reder’s (2002) Experiment 2, and so there was less mixing for low-frequency new words in Arndt and Reder’s study; this might be due to differences in the low-frequency words used across the studies.

A reviewer asked whether the LN* distribution can simply be restricted to have the same location as the LO distribution, as the LN* words are viewed as being low-frequency new words that are treated like LO words. The parameter estimates and standard errors in Table 4 show that this does not hold for Experiments 1a and 1b (but possibly for Arndt and Reder’s 2002 Experiment 2), whereas Table 7 (later) shows that this also does not hold for Arndt and Reder’s (2002) Experiment 1. Thus, the results suggest that there are differences between low-frequency new words that are considered as old (LN* words) and low-frequency words that are actually old (LO words), as reflected by the different levels of familiarity—that is, the LN* words do not appear to be as familiar as actual LO words.

In summary, an application of mixture SDT models to mirror effect data yields several interesting results. First and most important, the experiments show clear and consistent differences in the mixing parameter across low- and high-frequency words, in that the estimate of λ_{LO} was larger than the estimate of λ_{HO} in all three experiments. Thus, the mixing parameter systematically varies with word frequency, which is consistent with results found in earlier research; this result can be interpreted as showing that low-frequency words receive more attention or are processed more deeply than high-frequency words, exactly as speculated by many researchers. Second, a reversal in the order of the LO and HO distributions was not found in any of the studies. More specifically, the results for Experiments 1a and 1b suggest that the LO and HO distributions have the same location, and those for Arndt and Reder’s (2002) Experiment 2 suggest that LO and HO have either the same location or HO is to the right of LO. In any case, a reversal in strength did not appear in any of the experiments once effects of mixing were controlled for. The results suggest that the apparent reversal found for a conventional SDT analysis occurs because of the different levels of attention across high- and low-frequency words, which are confounded with effects of strength.

Individual Data

The results examined above were for pooled data, which may or may not be reflective of individual data. One can also fit the Mixture 2 model to individual data, but it should be recognized that there are some limitations, namely that a sample size of 280 for each individual, though typical, is somewhat small; the small sample size leads to problems with estimation, mainly that the model is in many cases empirically not identified when applied to individual data (see DeCarlo, 2003a, p. 776). This does not indicate a problem with the model per se but basically means that larger sample sizes are needed in order to apply the model to individual data. If interest centers on individual data, then one possibility is to increase the session length and obtain more responses, which will alleviate the identification problem. Another possibility is to use a multilevel version of the mixture SDT model (see DeCarlo, 1998, p. 197); this approach allows for individual differences and uses information across participants as well as within; the approach merits investigation in future research.

To obtain information about results for individual data, the Mixture 2 model was fit to the data of each participant in both Experiments 1a and 1b. In one case, the program failed to converge; in many other cases problems with empirical identification arose (and the standard errors were indeterminate). However, indeterminate standard errors did not appear for 27 cases in Experiment 1a and for 12 cases in Experiment 1b. Table 5 shows the mean parameter estimates for fits of the mixture model to the individual data of Experiments 1a and 1b (means weighted by the inverse of the square of the standard errors were also examined but gave similar results).[5]

The most important result shown in Table 5 is that the mixing parameter for low-frequency words is again larger than that for high-frequency words in both experiments, as found above for the pooled analysis. Second, there is again no evidence of a reversal in HO and LO distributions, in that the ds for the HO and LO distributions are about equal (again with a difference of 0.20 or

[5] In a few cases, there were problems determining the direction of labeling for the latent variables; see DeCarlo (2005) for some comments on label switching.

less), exactly as found above (there was considerable variability across participants, and so the standard deviations of the parameter estimates are large). Finally, Table 5 shows that the order of the LN, LO′, HN, HO′, and LN* distributions is again exactly the same as that found in Table 4. In short, Table 5 shows that results for the individual analyses are consistent with those found for the pooled data.

On Disentangling Effects of Strength From Those of Attention

From the perspective of mixture SDT, the above results show that if the mixing parameter differs across low- and high-frequency words, then a conventional SDT analysis will confound effects of strength with those of attention. This is illustrated in Figure 4. The top part of the figure shows new, partially attended, and old distributions for low- and high-frequency words (the figure shows an example in which HO is to the right of LO; note that the order of the distributions is the same as that found in Table 4). The arrows above the distributions indicate which distributions are mixed, and the numbers show the mixing proportions, using values similar to those found here. With respect to LO and HO words, the figure shows that 75% of the low-frequency words come from the LO distribution and 25% from the LO′ distribution, whereas 30% of the high-frequency old words come from the HO distribution and 70% from the HO′ distribution; that is, 75% of the low-frequency words are fully attended to, whereas only 30% of the high-frequency words are fully attended to. The lower part of Figure 4 shows the resulting mixed distributions (densities for actual mixtures of normal distributions are shown). The figure shows that, because of the higher attention (75%) to low-frequency words, the mixed LO distribution is not shifted as far to the left as

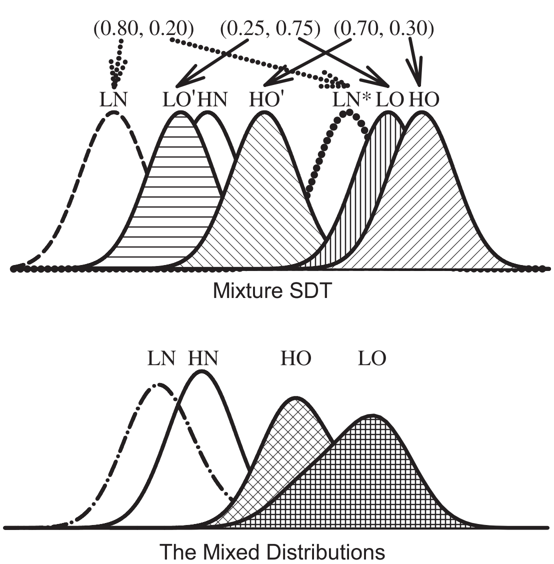

Figure 4. The top panel shows normal underlying distributions for the mixture signal detection theory (SDT) model applied to mirror effect data. The bottom panel shows the resulting mixed distributions. LN = low-frequency new words; HN = high-frequency new words; HO = high-frequency old words; LO = low-frequency old words; LN* = similar low-frequency new words; O′ = the partially attended distribution.

the mixed HO distribution, with the result that the mixed LO distribution is to the right of the mixed HO distribution (and has larger variance), even though the opposite is true for the unmixed distributions. Thus, comparing the top and lower panels in Figure 4, one sees a reversal in the order of the low- and high-frequency distributions across new and old words (i.e., the order in the top panel is LN, HN, LO, HO, whereas in the bottom panel it is LN, HN, HO, LO).

Note that the reversal in order shown from top to bottom in Figure 4 occurs only if there are differences in the mixing parameters across low- and high-frequency words. If the mixing parameters were about equal in magnitude across the two types of words, then a reversal might not occur (it also depends on the locations of the LO′ and HO′ distributions), in which case the order of the underlying distributions would be the same across a mixture SDT analysis and a conventional SDT analysis. Thus, in other situations where the mirror effect has been found (i.e., for variables other than word frequency), one cannot determine whether the effect is due to an actual reversal in location or is due to differences in attention; to help determine this, the effects of attention have to be controlled for, as done in a mixture SDT analysis.

A New Mixing Process: Effects of Similarity?

The mixture SDT model of Equation 3 introduces a new process, which is mixing of low-frequency new words, as suggested by the considerations discussed above. The interpretation offered here is that the mixing occurs because some low-frequency new words are correctly recognized as new (LN words), but others (LN* words) are treated like old words, possibly because their meaning is not known or because of surface similarities to LO words. This interpretation is supported by the finding in Table 4 that the LN* distribution was clearly to the right of the LN distribution in Experiment 1a, Experiment 1b, and Arndt and Reder's (2002) Experiment 2. This is also illustrated in Figure 4. The top panel shows that the unmixed LN distribution is to the left of HN distribution, whereas the LN* distribution is close to the old word distributions. The lower panel of Figure 4 shows that mixing shifts the mixed LN distribution to the right, though not far enough to cross over the HN distribution, in that the mixed LN distribution is still to the left of the HN distribution, as found in the three experiments presented above. This occurs because the unmixed LN distribution is far to the left of the HN distribution and because the amount of mixing is relatively small (e.g., 30% in Experiment 1a, 34% in Experiment 1b, and only 4% in Arndt & Reder's study), and so the amount that the mixed LN distribution is shifted is small.

The results for LN words illustrated in Figure 4 suggest an interesting prediction: If the mixing was increased, then the LN and HN distributions might appear, from the conventional SDT perspective, to reverse positions; that is, with sufficient mixing, the mixed LN distribution might appear to cross over the HN distribution (exactly as for the LO and HO distributions) and therefore be to the right of it. Fortunately, data to test this prediction were available. In particular, an experiment of Arndt and Reder (2002, Experiment 1) used words that were purposely constructed to be similar to old words. This was done by using, as new words, plurality-reversed versions of some of the old words (e.g., if *home* was an old word, a similar but new plurality-reversed word would be *homes*), as earlier done by Hintzman and Curran (1994).

Compared with LN words, it seems that a larger percentage of plurality-reversed new words should be mistaken for old words, because they were purposely constructed to be similar to old words (they are in fact simply old words with reversed plurality). Thus, a basic prediction is that the plurality-reversed new words (hereafter denoted simply as LS and HS words, for *low-similar* and *high-similar* new words) should have larger mixing parameters than the LN words. Second, the unmixed distributions should behave like the LN and LN* distributions found above; that is, the LS* and HS* distributions should be (far) to the right of the LS and HS distributions, because the LS* and HS* distributions represent falsely recognized new words, whereas the LS and HS distributions represent new words that are correctly recognized as new. Third, if the mixing is great enough, then from the conventional SDT perspective, the mixed LS and HS distributions might appear to cross over the LN and HN distributions, respectively, and so the distributions will reverse locations across a mixture analysis and an unequal variance analysis. Note that this result did not appear for LN words in the three experiments presented above (i.e., the mixed LN distribution did not cross over the HN distribution, in that the LN distribution was to the left of the HN distribution in both mixture and unequal variance SDT analyses) because the amount of mixing for LN words was relatively small.

Arndt and Reder (2002) Experiment 1

Method

Participants and Design

The study consisted of 35 students who were each presented with a study list of 80 words (plus 4 buffers), which consisted of 40 low-frequency words and 40 high-frequency words (the words were the same as those used in Arndt and Reder's [2002] Experiment 2, described above) presented for 2 s each. Half of each type of old word were in singular form and half were in plural form. During the test, participants were presented with 120 words that consisted of 40 studied words (half low frequency and half high frequency), 40 new words (again half of each frequency), and 40 similar new words (again half of each frequency). The similar new words were plurality-reversed versions of the old words, which were to be treated as new words. So, for example, if the word *thoughts* appeared on the study list and *thought* on the test list, then *thought* should be treated as new. Clearly, plurality-reversed new words are highly similar to old words, which is why they were used. Participants were run in three study–test cycles, giving a total of 360 trials per participant. Further details can be found in Arndt and Reder (2002).

The Mixture Model

The mixture model is the same as that given in Equation 3—that is, the Mixture 2 SDT model—except that it is extended to include terms for low- and high-frequency plurality-reversed new words (i.e., LS and HS words). As for LN words in Equation 3, the Mixture 2 SDT model allows for mixing for LS and HS words, as it seems highly likely that some of the plurality-reversed new words were treated as old words (i.e., were falsely recognized), considering that they were simply plurality-reversed versions of some of the old words.

Table 6
Goodness-of-Fit Statistics and Information Criteria for Experiment 1 of Arndt and Reder (2002)

Model	LR	*df*	*p*	AIC	BIC
Unequal variance SDT	94.42	15	<.001	41,422	41,453
Mixture SDT (Mix 2)	10.46	10	.401	41,348	41,390
Dual-process model	652.03	18	<.001	41,973	42,062

Note. N = 12,600. LR = likelihood ratio goodness-of-fit test; AIC = Akaike information criterion; BIC = Bayesian information criterion; SDT = signal detection theory; Mix 2 = Equation 3.

Results and Discussion

Goodness of Fit

Table 6 shows LR goodness-of-fit statistics and information criteria for both the unequal variance model and the Mixture 2 model. The LR statistics are large and significant for the unequal variance SDT model, which indicates poor fit. Thus, Arndt and Reder's (2002) Experiment 1 joins the three experiments discussed above in showing a lack of fit for the unequal variance SDT model applied to mirror effect data. The Mixture 2 SDT model, on the other hand, gives a nonsignificant LR statistic, and so it adequately describes the data. The Mixture 2 SDT model is also clearly favored by both AIC and BIC, in that they are considerably smaller in value.

Figure 5 shows a z-ROC plot of the data and fitted z-ROC curves for the Mixture 2 model. It is evident from the plot why the unequal variance SDT model does not fit, in that the z-ROC curves clearly show nonlinearity, as was also found for the experiments presented above. For example, the LN (open circles) and HO (gray circles) data show upward curvature, exactly as found for the three experiments discussed above, and the LO (black circles) data show a dip, also as found for Experiments 1a and 1b. A new result is that the z-ROC curves for the plurality-reversed words (i.e., LS and HS words) are also clearly nonlinear, particularly for LS words; this result suggests mixing. Overall, the figure shows that z-ROC curves for the mixture SDT model accurately describe the data.

Parameter Estimates

Table 7 shows parameter estimates for the fitted models. The top part of the table shows parameter estimates for fits of the unequal variance SDT model. The mirror effect again appears, in that the distributions are ordered as LN, HN, HO, LO, and given the small standard errors, it is clear that the LO distribution is to the right of the HO distribution and the LN distribution is to the left of the HN distribution. With respect to the variances, the variances of the LN, HN, LO, and HO distributions are ordered the same as in Table 3, with HN words again having the smallest variance. There is, however, not a clear pattern to the variances when LS and HS words are considered. For example, LS words have variance as high as LO words, even though the LS word distribution has the same location as the HN words, and so the magnitude of the variance does not vary systematically with location.

An interesting result for the unequal variance SDT model is that the false alarm rates are higher for the LS and HS words as compared with LN and HN words; that is, the LS and HS distributions are to the right of the LN and HN distributions. As

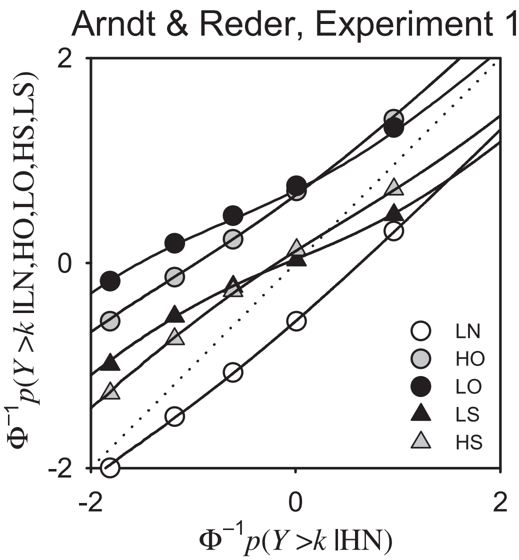

Figure 5. Data and fitted *z*-transformed receiver operating characteristic (*z*-ROC) curves for Experiment 1 of Arndt and Reder (2002). LN = low-frequency new words; HN = high-frequency new words; HO = high-frequency old words; LO = low-frequency old words; LS = low-similar new words; HS = high-similar new words.

discussed above, this result follows from mixture SDT if there is great enough mixing for LS and HS words, as examined next.

The lower part of Table 7 shows parameter estimates for fits of the Mixture 2 SDT model. The LN distribution is again to the left of the HN distribution, as in the mirror effect, and the point estimate for the LO distribution (2.24) is slightly larger than that for the HO distribution (2.02). However, the difference (0.22) is small in magnitude, and taken with the standard errors, there does not appear to be evidence of a difference in the locations of the HO and LO distributions; an LR test of $d_{HO} = d_{LO}$ gives a test statistic of 1.90 with $df = 1$ and $p = .17$, and so the null hypothesis of equal locations is not rejected. Thus, Table 7 shows that once effects of mixing were controlled for, there was no evidence of any difference in the location of the LO and HO distributions, as was found in the three experiments discussed above. In contrast, the results for the unequal variance SDT model clearly indicated a reversal of the LO and HO distributions.

Table 7 also shows that the proportion of attended words is higher for LO words (estimate of $\lambda_{LO} = 0.64$) than for HO words (estimate of $\lambda_{HO} = 0.46$), which suggests that attention is greater for LO words than for HO words, as found for the three experiments presented above. It is also apparent that the LN* distribution is far to the right of the LN distribution (1.73), which is consistent with the view that it represents falsely recognized new words. The table also shows that the proportion of LN* words is small, .10, and so only 10% of the low-frequency new words are confused with old words, which is consistent with results found for Arndt and Reder's (2002) Experiment 2 (where the estimated proportion of LN* words was .04). Note that the unmixed LN distribution is still to the left of the unmixed HN distribution, which follows because the amount of mixing is small, as discussed above.

An important result shown in Table 7 is that the estimated mixing parameters are clearly larger for LS* and HS* words ($\lambda_{LS} = 0.51$ and $\lambda_{HS} = 0.62$) than for LN* words ($\lambda_{LN*} = 0.10$). Thus, LS and HS words were falsely recognized much more often (over 50% of the time) than LN words (10% of the time); this is yet more evidence showing that the mixing parameter (λ) behaves as expected in response to experimental manipulations (i.e., it is larger when similar new words are used). The larger values of the mixing parameters found for LS and HS words also explain why, when analyzed using conventional SDT, the LS and HS distributions appear to be to the right of the LN and HN distributions—because of the greater mixing, as discussed above. Also note that the LS* and HS* distributions are far to the right of the LS and HS distributions, which is consistent with the view that the plurality-reversed words were sometimes correctly treated as new (LS and HS) and at other times were incorrectly treated as old (LS* and HS*); it is also consistent with the view that the same thing occurs for LN words in mirror effect experiments.

Overall, the results for Experiment 1 of Arndt and Reder (2002) suggest that there was mixing for plurality-reversed words. Most

Table 7
Parameter Estimates for Mixture and Unequal Variance Signal Detection Models for Experiment 1 of Arndt and Reder (2002)

	Unequal variance SDT model					
Parameter	LN	HN	LS	HS	HO	LO
d'	−0.62 (0.04)	0	0.02 (0.05)	0.12 (0.04)	1.00 (0.04)	1.50 (0.06)
log σ_s	0.16 (0.03)	0	0.67 (0.03)	0.36 (0.03)	0.37 (0.03)	0.65 (0.03)

	Mixture 2 model										
	LS	HS	LN	LO′	HN	HO′	HS*	LN*	LS*	HO	LO
d'	−1.30 (0.08)	−1.14 (0.16)	−0.76 (0.07)	−0.32 (0.08)	0	0.12 (0.10)	0.86 (0.09)	0.97 (0.31)	1.38 (0.07)	2.02 (0.16)	2.24 (0.08)
		$\lambda_{LO} = 0.64$ (0.02)		$\lambda_{HO} = 0.46$ (0.05)		$\lambda_{LN*} = 0.10$ (0.04)		$\lambda_{LS} = 0.51$ (0.02)		$\lambda_{HS} = 0.62$ (0.05)	

Note. The first row for each experiment shows estimates of *d*, with standard errors in parentheses. SDT = signal detection theory; LN = low-frequency new words; HN = high-frequency new words; LS = low-similar new words; HS = high-similar new words; HO = high-frequency old words; LO = low-frequency old words; LO′ = unattended low-frequency old words; HO′ = unattended high-frequency old words; HS* = high-similar new words treated as old; LN* = low-frequency new words treated as old; LS* = low-similar new words treated as old; λ_{LO} = low-frequency attention; λ_{HO} = high-frequency attention; λ_{LN*} = low-frequency new mixing; λ_{LS} = low-frequency similar mixing; λ_{HS} = high-frequency similar mixing.

important, the results showed that the amount of mixing was considerably larger for LS and HS words (>50%) as compared with LN words (10%). This supports the interpretation that mixing occurs because of similarity. Second, as a consequence of thegreater mixing, the mixed LS and HS distributions "crossed over" the LN and HN distributions, as shown by fits of the unequal variance SDT model, exactly as predicted above. That is, the order of the distributions for the mixture SDT analysis was LS, HS, LN, HN, but for the unequal variance SDT analysis it was LN, HN, LS, HS; the mixture SDT analysis suggests that this difference occurred because of the greater mixing for LS and HS words. Finally, the amount of mixing for LN words was very small (10%), as also found in Arndt and Reder's Experiment 2 (4%), and so the mixed LN distribution did not cross over the HN distribution (i.e., it was to the left of it), as was also found in Experiments 1a and 1b and in Arndt and Reder's Experiment 2 above.

A reviewer wondered why, in the mixture SDT analysis, the LS and HS words were to the left of the LN and HN words. This finding suggests that LS and HS words were more likely to be recognized as new than LN and HN words. It should be noted that LS and HS words, though new, differ in a basic way from actual new words (i.e., LN and HN words) in that they are necessarily tied to old words because they are simply old words that are reversed in plurality. Thus, additional information is available for LS and HS words; this information likely helps one to recognize that they are new words, thereby giving LS and HS distributions that are further to the left. For example, if *home* is a familiar (old) word, then the participant can be fairly sure that *homes* is a new word, whereas this type of information is not available for LN and HN words. This is similar to a recall-to-reject account (e.g., Clark, 1992), with the difference that the word does not have to be recalled but can simply be very familiar; see the section on the dual-process model below (note that the negative values of d found for LS and HS words are not large enough to suggest a recall-to-reject interpretation).

Related Findings

The finding of mixing on new word trials, and the interpretation that this occurs because of similarity, is also consistent with other results found in the literature. For example, Malmberg et al. (2004) also constructed new words by using plurality-reversed words and found, from the conventional SDT perspective, that false alarm rates were higher for the plurality-reversed words. As just discussed, this result is consistent with greater mixing for plurality-reversed words. Park et al. (2005) also found higher false alarm rates for plurality-reversed new words. However, an interesting aspect of their results was that higher false alarm rates were not found when participants were informed about the features to be discriminated. Here I note that this result has a simple interpretation in terms of mixture SDT—it seems likely that the "informed condition" led participants to more deeply process the words; in fact, participants in this condition were told to generate an image of each word. If this led to deeper processing or greater attention, then there would be less mixing for plurality-reversed words (i.e., fewer were confused with old words), with the result that the mixed distribution would not be shifted as far to the right, and so a higher false alarm rate would not be found (depending on whether the amount of mixing was sufficiently lower).

In short, Park et al.'s (2005) results join Malmberg et al.'s (2004) results in that they both suggest the presence of a mixing process for plurality-reversed new words; an interesting aspect of Park et al.'s results is that they also suggest that higher false alarm rates might not have been found in the "informed condition" because the condition led to deeper encoding of plurality-reversed words, which resulted in less mixing, and so the distributions did not cross over. The use of a mixture SDT analysis might help to shed light on this and other manipulations used in other studies.

General Discussion

Are low-frequency words more memorable than high-frequency words? From the perspective of conventional SDT, the order of the hit and false alarm rates indicates the order of the underlying distributions, and so the answer is yes, because the LO distribution was clearly to the right of the HO distribution in all of the experiments examined here, and so low-frequency words are more memorable. One must realize, however, that the unequal variance SDT model is only one possible extension of the simple equal variance SDT model, and it is an extension that even Green and Swets (1988) acknowledged is not theoretically motivated (see DeCarlo, 2002, in which other problems with the model are noted). In contrast, the mixture SDT model has a simple theoretical motivation, in that many processes suspected of operating in recognition memory quite naturally lead to latent classes, and this in turn leads to mixing. For example, different levels of attention to the study words can lead to mixing because some of the old words then actually consist of two latent classes of old words, such as attended and partially attended words. In that case, the order of hits and false alarms reflects the order of the mixed distributions, and not the order of the unmixed distributions, as shown in Figure 4.

The Mirror Effect and Mixture SDT

The mixture SDT model suggests that the mirror effect for word frequency occurs because of differential attention to, or processing of, high- and low-frequency words. As shown in Figure 4, mixing can make it appear as if the (mixed) HO distribution is below the LO distribution, even if it is actually at or above the LO distribution. The results for Experiment 1a with native English speakers, Experiment 1b with nonnative English speakers, Arndt and Reder's (2002) Experiment 2, and Arndt and Reder's Experiment 1 all support this interpretation: The mixture parameter λ was larger for LO words than for HO words in all four experiments, and once the effects of mixing were controlled for, there was no apparent difference in the location of the LO and HO distributions. Thus, the results call into question whether low-frequency words are more memorable than high-frequency words. In fact, the four experiments examined here simply suggest that although high-frequency words were initially more familiar than low-frequency words (as expected and as shown by the HN and LN distributions), the words were about equally familiar after being studied (as shown by the unmixed HO and LO distributions).

Similarity and Mixing

A new result is that another mixing process might be operating for new words. An interpretation offered here is that the mixing reflects a tendency to treat some LN words as old, possibly because of similarity to LO words (e.g., LN and LO words are both more unusual, as compared with high-frequency words). This not only provides an account of the nonunit slopes and nonlinearity that is evident in the z-ROC curves for LN words but also is supported by an experiment of Arndt and Reder (2002), in which purposely constructed similar new words were used. The results in that case showed that the unmixed plurality-reversed words (LS* and HS*) behaved exactly like LN* words. The mixture SDT model can also account for the higher false alarm rates and larger variance found for plurality-reversed new words when the unequal variance SDT model is used, in that both of these results are basic consequences of mixing, as discussed above.

In addition to accounting for and providing new insights into results found with low-frequency words and plurality-reversed words, the Mixture 2 SDT model is potentially informative with respect to results found in other experiments, such as with rare words. In particular, several researchers have found that rare words differ from low-frequency words in that they tend to have higher false alarm rates (Rao & Proctor, 1984; see Wixted, 1992). In light of the results presented above, the higher false alarm rates could simply reflect greater mixing for rare words than for low-frequency words (perhaps because fewer are understood, as speculated by Wixted, 1992); greater mixing will shift the (mixed) distribution for rare words further to the right, resulting in higher false alarm rates. A mixture SDT analysis might also help to explain why some inconsistent results have been found across studies with rare words (see Wixted, 1992)—there might have been differences in encoding as well as strength across the studies, which are confounded in a conventional SDT analysis.

The Dual-Process Model

Although the focus here is on signal detection models, another model that has received attention in the memory literature is the dual-process model (e.g., Yonelinas, 1994, 1999). The dual processes being referred to are familiarity, as in conventional SDT, and recollection. Arndt and Reder (2002), for example, considered a dual-process model for their data, where parameters for recollection of low- and high-frequency old items were included, with the restriction that for recollected items, the response was simply the highest response category (which was 6). For the basic mirror effect experiment, the model can be written as four equations for the four types of stimuli (i.e., HO, LO, LN, and HN words, respectively).

$$\begin{aligned} p(Y > k|\mathrm{HO}) &= R_{\mathrm{HO}} + (1 - R_{\mathrm{HO}})\Phi(-c_k + d_{\mathrm{HO}}), \\ p(Y > k|\mathrm{LO}) &= R_{\mathrm{LO}} + (1 - R_{\mathrm{LO}})\Phi(-c_k + d_{\mathrm{LO}}), \\ p(Y > k|\mathrm{LN}) &= \Phi(-c_k + d_{\mathrm{LN}}), \\ p(Y > k|\mathrm{HN}) &= \Phi(-c_k), \end{aligned} \tag{5}$$

where R_{HO} and R_{LO} are the proportion of recollected HO and LO words, respectively, and the highest response category is used for recollected words. Equation 5 is the version of the dual-process model that has been used for mirror effect data (e.g., Arndt & Reder, 2002) as well as for recognition data (cf. Yonelinas, 1999, Equations 1 and 2).[6]

The tables presented above show fit statistics for fits of the dual-process model to the current data. Table 2 shows that the LR fit statistics are large and significant for Experiments 1a and 1b and for Arndt and Reder's (2002) Experiment 2, whereas Table 4 shows a large significant LR statistic for Arndt and Reder's Experiment 1. Thus, like the unequal variance SDT model, the dual-process model fails to describe the data. With respect to relative fit, the AIC statistics in Tables 2 and 4 clearly favor the Mixture 2 SDT model over the dual-process model; BIC also generally favors the mixture SDT model, with the exception of Arndt and Reder's Experiment 2.

Why has the lack of fit of the dual-process model to mirror effect data not been noted until now? There are two reasons, one being that fit statistics have previously not been reported and the second having to do with the way ROC plots have been presented. First, it should be noted that if one has four types of stimuli, as in mirror effect experiments (i.e., LN, HN, LO, and HO words), then one must consider results for the four types of stimuli simultaneously, that is, in one comprehensive model, as in the four equations of Equation 5. Previous research, on the other hand, has tended to consider the components of the model in a piecemeal manner; for example, there has been a tendency to fit each equation separately (in a spreadsheet) rather than fitting all four equations simultaneously and assessing fit. Similarly, ROC curves for mirror effect data should include three curves in one plot (one distribution is used as the reference), as in Figure 3, and not separate plots for different curves, as has been done in prior research. As shown in Figures 3 and 5, a plot of all of the curves together reveals nonlinearity in some of the curves.

The figures and Equation 5 also show why the dual-process model does not fit: The dual-process model predicts unit slope linear z-ROC curves for LN words, which was not found (the curves do not have unit slopes and are also nonlinear); this has been overlooked up to this point (because, for example, researchers have not plotted LN words against HN words). It should be noted that one can generalize the dual-process model and improve fit by allowing for unequal variances for LN words (i.e., nonunit slopes); however, a decision must then be made as to whether to do this for LO and HO words as well. In addition, it does not appear that this will totally solve the problem because, as shown in the figures, the unequal variance SDT model does not describe the nonlinearities found in the z-ROC plots (for LN and other word types), and so neither will a dual-process model with unequal variances. At the least, proponents of the dual-process model need to examine possible extensions of the model.

It should be noted that the Mixture 1 SDT model is also a "dual-process" model, in that it incorporates two processes—familiarity and attention. The Mixture 2 model is a triple-process model, in that it includes processes that result from effects of familiarity, attention, and similarity. Some insights into the dual-process model can be gained by noting that it can be viewed as a special case of the Mixture 1 SDT model. First, note that the dual-process model of Equation 5 can be written in terms of cumulative probabilities, as done for all of the models here, as follows:

[6] Note that Yonelinas (1999) parameterized the model using effect coding, whereas dummy coding was used here (see DeCarlo, 1998, p. 201).

$$p(Y \le k|\mathrm{HO}) = (1 - R_{\mathrm{HO}})\Phi(c_k - d_{\mathrm{HO}}),$$
$$p(Y \le k|\mathrm{LO}) = (1 - R_{\mathrm{LO}})\Phi(c_k - d_{\mathrm{LO}}),$$
$$p(Y \le k|\mathrm{LN}) = \Phi(c_k - d_{\mathrm{LN}}),$$
$$p(Y \le k|\mathrm{HN}) = \Phi(c_k). \quad (6)$$

Next, it is simple to show that Equations 1 and 2 for the Mixture 1 SDT model give the following four equations for HO, LO, LN, and HN words:

$$p(Y \le k|\mathrm{HO}) = \lambda_{\mathrm{HO}}\Phi(c_k - d_{\mathrm{HO}}) + (1 - \lambda_{\mathrm{HO}})\Phi(c_k - d_{\mathrm{HO}'}),$$
$$p(Y \le k|\mathrm{LO}) = \lambda_{\mathrm{LO}}\Phi(c_k - d_{\mathrm{LO}}) + (1 - \lambda_{\mathrm{LO}})\Phi(c_k - d_{\mathrm{LO}'}),$$
$$p(Y \le k|\mathrm{LN}) = \Phi(c_k - d_{\mathrm{LN}}),$$
$$p(Y \le k|\mathrm{HN}) = \Phi(c_k). \quad (7)$$

A comparison of Equation 7 with Equation 6 shows that the only differences are the terms $\lambda_{\mathrm{HO}}\ \Phi(c_k - d_{\mathrm{HO}})$ and $\lambda_{\mathrm{LO}}\ \Phi(c_k - d_{\mathrm{LO}})$ in the first two equations. If these terms are set to zero, then Equation 7 reduces to the dual-process model of Equation 6, with λ_{HO} and λ_{LO} becoming R_{HO} and R_{LO} (and $d_{\mathrm{HO}'}$ and $d_{\mathrm{LO}'}$ of Equation 7 becoming d_{HO} and d_{LO} of Equation 6); note that these terms can be set arbitrarily close to zero by fixing d_{HO} and d_{LO} to large values. Thus, Equation 7 shows that the dual-process model is related to a special case of the mixture SDT model; that is, the Mixture 1 SDT model essentially reduces to the dual-process model when d_{LO} and d_{HO} are restricted to large values.[7] A practical consequence of this observation is that the dual-process model can be approximated to any desired level of accuracy by restricting d_{LO} and d_{HO} in the mixture model to large values; this suggests a simple way to implement maximum likelihood estimation for the dual-process model through the use of software for mixture models.

Figure 6 illustrates how "recollection" can be conceptualized and implemented within the framework of mixture SDT. The two components of the mixture can be interpreted as one based on familiarity (the O and N distributions) and one related to recollection (the $\mathrm{O_R}$ distribution). The figure shows that if the distribution for one component of the mixture model is restricted to be far to the right (the $\mathrm{O_R}$ distribution), then the results are consistent with those interpreted as indicating recollection. That is, it should be apparent from Figure 6 that the probability of a response of "4" (or whatever the highest response category is) can be made arbitrarily close to 1 by restricting *d* to a large value for the "recollected" distribution (with the result that the probability of responses of 1 to 3 are arbitrarily close to zero). For example, for the four experiments examined here, the probability of a response of 6 was greater than 0.99999 for values of *d* of 8 or greater (and the results did not change to three decimal places when larger values were used). This basically means that one cannot distinguish between high familiarity (i.e., a probability of the highest response category close to unity) and recollection (a probability of the highest response category equal to unity); in both cases, participants are quite sure that they have seen the word. In short, in a signal detection framework, recollection can be conceptualized simply as high familiarity. Note that the above results show that a fit of the mixture SDT model, which does not place restrictions on the locations of the distributions, gives (unmixed) rightmost distributions that are not that far to the right (i.e., the estimates of *d* are relatively small), and in this way the unmixed distributions do not appear to represent "recollection." A similar point was made above with respect to the LS and HS distributions—they are not far enough to the left to support a recall-to-reject account.

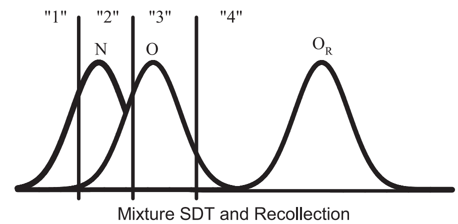

Figure 6. An illustration of how recall is handled in signal detection theory (SDT) as high discrimination, with N and O indicating the new and old word distributions based on familiarity and $\mathrm{O_R}$ indicating the old word distribution based on "recollection."

In summary, like the unequal variance SDT model, the current version of the dual-process model fails to describe mirror effect data. The dual-process model can also be viewed as being related to a special case of the Mixture 1 SDT model, albeit with a different interpretation. The relation shows that recollection can be viewed in the framework of SDT simply as a high level of familiarity. This in turn suggests a simple way to fit the dual-process model using standard software for mixture analysis.

Conclusions

The purpose here has been to show that within the framework of SDT, there is more than one way to look at the data. It appears to be worthwhile to consider mixture SDT models, in that the results can differ considerably once effects of mixing are taken into account. The mirror effect for word frequency is a good example, in that the underlying distributions do not show a reversal across new and old words once effects of mixing have been taken into account. It is shown that mirror effect data present challenges to both the unequal variance SDT model and the currently popular dual-process model, in that neither model describes the data, whereas the extended mixture SDT model takes some tentative steps toward dealing with these challenges. The mixture SDT model also unifies a variety of results, such as those obtained with high- and low-frequency words, with plurality-reversed words, and with rare words. The mixture SDT approach offers a new perspective and suggests new research possibilities. The results are compelling and suggest that further applications of mixture SDT models might lead to a deeper understanding of processes operating in recognition memory. The model-based approach offered here also provides a detailed and unified way of comparing different theories across studies.

[7] More precisely, the dual-process model is a mixture model with different types of models across the two components of the mixture. Here it is being approximated by using the same type of model across both components, as in mixture SDT.

References

Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 830–842.

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31,* 297–305.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20,* 231–243.

Dayton, C. M. (1998). *Latent class scaling analysis.* Thousand Oaks, CA: Sage.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3,* 186–205.

DeCarlo, L. T. (2000, August). *An extension of signal detection theory via finite mixtures of generalized linear models.* Paper presented at the 2000 meeting of the Society for Mathematical Psychology, Kingston, Ontario, Canada.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109,* 710–721.

DeCarlo, L. T. (2003a). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 767–778.

DeCarlo, L. T. (2003b). Using the PLUM procedure of SPSS to fit unequal variance and generalized signal detection models. *Behavior Research Methods, Instruments, & Computers, 35,* 49–56.

DeCarlo, L. T. (2005). A model of rater behavior in essay grading based on signal detection theory. *Journal of Educational Measurement, 42,* 53–76.

Dobbins, I. G., & Kroll, N. E. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1186–1198.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics* (Rev. ed.). Los Altos, CA: Peninsula.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75–102). New York: Wiley.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hintzman, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 201–205.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33,* 1–18.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302–313.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Lillard, L. A., & Panis, C. W. A. (2003). aML Multilevel Multiprocess Statistical Software (Version 2.0). Los Angeles: EconWare.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. (1976). Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 75–102). New York: Wiley.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 539–559.

Malmberg, K. J., Holden, J., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 319–331.

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition, 31,* 35–43.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760.

Murdock, B. B. (1998). The mirror effect and attention-likelihood theory: A reflective analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 524–534.

Murdock, B. B. (2003). The mirror effect and the spacing effect. *Psychonomic Bulletin & Review, 10,* 570–588.

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide* (4th ed.). Los Angeles: Author.

Overschelde, J. P. V. (2002). The influence of word frequency on recency effects in directed free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 611–615.

Park, H., Reder, L. M., & Dickison, D. (2005). The effects of word frequency and similarity on recognition judgments: The role of recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 568–578.

Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 386–394.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 763–785.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294–310.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM— retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

Sikström, S. (2001). The variance theory of the mirror effect in recognition memory. *Psychonomic Bulletin & Review, 8,* 408–438.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1379–1396.

Vermunt, J. K. (1997). *LEM: A general program for the analysis of categorical data* [Computer software and manual]. Retrieved September 1998 from the Tilburg University Web site: http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 681–690.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354.

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 7, 2006
Revision received August 30, 2006
Accepted September 5, 2006 ■