Recognition memory for source and occurrence:
The importance of recollection

JOEL R. QUAMME
University of California, Davis, California

CHRISTINA FREDERICK
University of California, Berkeley, California

NEAL E. A. KROLL and ANDREW P. YONELINAS
University of California, Davis, California

and

IAN G. DOBBINS
Harvard University, Cambridge, Massachusetts

Previous recognition memory studies indicate that when both recollection and familiarity are expected to contribute to recognition performance (e.g., discriminating studied items from nonstudied items) the dual-process and the unequal-variance signal detection models provide comparable accounts of performance. When familiarity is not expected to be useful (e.g., when items from two equally familiar sources are discriminated between), the dual-process model provides a significantly better account of performance. In the present study, source recognition was tested under conditions in which familiarity could have been used to perform a list-discrimination task; participants were required to discriminate between strong studied items, weak studied items, and new items. The dual-process model provided a better account of performance than did the unequal-variance model. Moreover, the results indicated that the unequal-variance assumption in a single-process signal detection model was not a valid substitution for recollection and that recollection was used to make recognition judgments even when assessments of familiarity were useful.

Many classic and current models of recognition memory assume that old/new recognition decisions (i.e., "was this item presented before?") are based on an assessment of unidimensional memory strength, or *familiarity* (e.g., Banks, 1970; Donaldson, 1996; Green & Swets, 1966; Hirshman & Master, 1997; Inoue & Bellezza, 1998). Test items are represented as falling along a familiarity continuum so that studied (i.e., old) items are more familiar, on average, than nonstudied (i.e., new) items. Most models incorporate signal detection theory by assuming that variability in the familiarity values of new and old items results in partially overlapping normal distributions. Because old and new distributions overlap, an individual must select a specific familiarity level (i.e., a response criterion) in order to decide whether or not a test item was presented previously. Items falling above the criterion are judged to be old, and items below the criterion are judged to be new. The hit rate is defined as the proportion of the old-item distribution that exceeds the response criterion, and the false alarm rate is defined as the proportion of the new-item distribution that exceeds the response criterion (see Macmillan & Creelman, 1991, for an overview of signal detection theory).

Dual-process models of recognition (e.g., Atkinson & Juola, 1974; Jacoby, 1991; Jacoby & Dallas, 1981; Mandler, 1980, 1991; Yonelinas, 1994) assume that a familiarity judgment by itself is not sufficient to explain how recognition judgments are made. These models assume that an additional process, the *recollection* of qualitative information, may also contribute to recognition judgments. In dual-process models, familiarity or strength has been conceptualized as continuous (Atkinson & Juola, 1974; Yonelinas, 1994), automatic (Jacoby, 1991; Jacoby & Dallas, 1981), and item specific (Mandler, 1980, 1991). Conversely, recollection is often referred to as a threshold process (Rotello, Macmillan, & Van Tassel, 2000; Yonelinas, 1997, 1999a), controlled (Atkinson & Juola, 1974; Jacoby, 1991) and accompanied by the retrieval of contextual associations (Gardiner, 1988; Jacoby, 1991; Tulving, 1985; Yonelinas, 1997, 1999a).

The aim of the present paper is to assess the ability of single- and dual-process models of recognition to account for source recognition receiver-operating characteristics

(ROCs) in conditions under which both models predict that familiarity should be useful for making recognition judgments. Participants were required to discriminate studied target items from both studied lures that differed in strength and new items. This type of task requires the simultaneous recognition of source and occurrence and provides unique constraints on how familiarity is used in the two models to make recognition decisions. Specifically, we asked whether recollection contributes to performance when it is unnecessary (i.e., when familiarity could be sufficient) and whether unequal variances of familiarity distributions are required if recollection is assumed to contribute. First, we will provide a brief overview of ROC curves in recognition. Then we will describe the use of familiarity-only and dual-process models to account for recognizing both occurrence and source.

The Use of Receiver-Operating Characteristics in Model Comparison

A direct method for assessing the predictions of models and the outcomes of experiments is the examination of ROCs. An ROC is an empirical function that relates hit rates to false alarm rates in a decision task, such as recognition. ROCs can be generated by having participants make confidence ratings for each recognition judgment. The confidence continuum often ranges from very confident *new* judgments (i.e., rejections) to very confident *old* judgments (i.e., endorsements). Hit rates are plotted against false alarm rates across levels of confidence, with the assumption that different levels of confidence result from more or less stringent criterion settings.

For a signal detection model of recognition with equal variances of old and new distributions, a curvilinear ROC is predicted that is symmetrical about the diagonal and always passes through the coordinates 0,0 and 1,1. When plotted in z-space, an equal-variance signal detection (EVSD) model predicts a linear function with a slope of 1.0. However, findings suggest that recognition memory ROCs are generally not symmetrical about the diagonal (Ratcliff, McKoon, & Tindall, 1994; Ratcliff, Sheu, & Gronlund, 1992). Specifically, in probability space, the ROC function tends to be "pushed" toward the y-axis, and in z-space, the slope is usually less than 1.0, suggesting that the variance of old items is greater than the variance of new items. Consequently, most unidimensional signal detection theories of recognition now incorporate an unequal-variance assumption, and some global-matching models explicitly predict greater old-item variance than new-item variance (e.g., Gillund & Shiffrin, 1984; Hintzman, 1986). Thus, an unequal-variance signal detection (UVSD) model can overcome one of the major deficiencies of the EVSD, in that it is compatible with the ROCs reported in most experiments.

Yonelinas (1994) developed the dual-process signal detection (DPSD) model, which combines a familiarity assessment based on signal detection theory with a threshold recollection process. The DPSD model also predicts ROCs that are curved and asymmetrical along the diagonal (Yonelinas, Dobbins, Szymanski, Dhaliwal, & King,

1996). However, according to this model, the asymmetry is caused by a disproportionate number of highly confident hits that represent recollected items. That is, familiarity is assumed to reflect an equal-variance signal detection process that alone produces a symmetrical ROC. However, some proportion of old items are expected to be recollected, which has the effect of increasing the hit rate and producing an asymmetrical ROC, like those observed in standard recognition tasks.

The major difference between UVSD and DPSD models is that the latter predicts an ROC with a nonzero intercept equal to a proportion of old items that exceed a recollection threshold. Unlike the signal detection process that assumes Gaussian memory strength distributions, the threshold recollection process assumes noncontinuous distributions for which there is some degree of nonoverlap between new and old distributions (i.e., the strengths of some old items exceed the highest possible strength of any new items; see Murdock, 1974, for a more complete treatment of threshold theory). A threshold process predicts an ROC with a nonzero intercept that is linear within the unit square, rather then curved (Murdock, 1974; Swets, 1986).[1] If familiarity (a signal detection process) and recollection (a threshold process) both contribute to performance, as the dual-process model assumes, the ROCs should be asymmetrical and curved, but they should be slightly more linear than those predicted by signal detection theory alone.

Contrasting the UVSD and the DPSD models by examining performance in standard recognition test procedures has been difficult because, in standard recognition tests, the DPSD model assumes that both recollection and familiarity contribute to performance and predicts ROCs that are almost identical to those predicted by the UVSD model. In some recognition studies, the DPSD model provides a small advantage over the UVSD model. In other studies, the pattern is reversed, and in many, the model fits are not significantly different (see Yonelinas, 1997, 1999a; Yonelinas et al., 1996; for further discussion, see Glanzer, Kim, Hilford, & Adams, 1999; Yonelinas, 1999b).

However, the two models do make different predictions under conditions in which the contribution of recollection is relatively large and that of familiarity is relatively small. Under these conditions, the DPSD model predicts that the ROC should not deviate much from linearity. For example, relatively linear ROCs have been found for recognition tasks in which individuals discriminate between intact and recombined arbitrary associations between studied words (Kelly & Wixted, 2001; Rotello et al., 2000; Yonelinas, 1997), among facial features (Yonelinas, Kroll, Dobbins, & Soltani, 1999), and between source or list contexts (Yonelinas, 1999a). These types of tasks require individuals to retrieve information about the studied association between two items or between an item and its context. That relatively linear ROCs are observed for many of these tasks is evidence that a continuous signal detection process does not contribute to the recognition of arbitrary associations. One interpretation of these results is that the familiarity of an item–item or

an item–context pairing is not increased beyond those of the individual items and contexts themselves (i.e., the constituent parts). That is, whereas item information becomes more familiar with study, there are cases in which associative and source information does not (see Kelly & Wixted, 2001, for a similar view).

Although a number of experiments have shown that the UVSD model accounts poorly for the shapes of source ROCs, a linear component is not always sufficient to account for the shape of the ROC. This has been demonstrated directly with source ROCs when items from two sources differ in strength (Yonelinas, 1999a, Experiment 4), when encoding procedures in the study phases are relatively rich (e.g., Qin, Raye, Johnson, & Mitchell, 2001), or when multidimensional rating scales are used to combine item recognition and source recognition in a single task (e.g., Slotnick, Klein, Dodson, & Shimamura, 2000). These findings suggest either that recollection is not always well described as a threshold process or that familiarity can sometimes be used to discriminate among items from different sources. Regardless of whether either or both of these possibilities hold, any adequate model must account for both linear and nonlinear ROCs and must provide empirically verifiable predictions of recollection and familiarity estimates across experimental conditions. To account for both linear and nonlinear ROCs, single-process signal detection models must assume that the threshold process does not contribute to performance under most circumstances (i.e., when ROCs are curvilinear) but that participants occasionally *switch* to a threshold process when continuous strength information is unhelpful. Whereas it is generally accepted that a threshold process of some sort is required to account for linear ROCs, it is less clear whether a threshold process such as recollection regularly contributes to performance when ROCs are nonlinear (Glanzer et al., 1999). It has been considered previously that although recollection may be used in some recognition tasks, it might not be used on recognition tests when it is not required or when it provides information redundant with that provided by familiarity (e.g., an inclusion task in the process dissociation procedure; Clark, 1999; Jacoby, 1991; Ratcliff, Van Zandt, & McKoon, 1995). A critical issue is whether it is reasonable to assume that recollection does not contribute to recognition when familiarity should be sufficient to discriminate between targets and lures. In other words, do individuals use both recollection and familiarity any time both processes provide useful mnemonic information about the study phase (as is assumed by dual-process models), or do they use recollection only when familiarity is insufficient (as is assumed by a single-process switching view)? To evaluate these positions, it is first necessary to define what conditions must be met for familiarity to support source discrimination.

Source Memory Discrimination Based on Familiarity

Here, we consider a simple form of familiarity-only recognition. For the present purposes, it was assumed that familiarity is a unidimensional signal detection process and that when familiarity is used in a recognition task, participants are able to control response criteria only, and not the actual computation of familiarity (see Clark, 1999, for a discussion of this assumption). For a familiarity-only signal detection model of this sort to account for list or source discrimination, it must be assumed that the familiarity distributions of items from the two lists have different means, different variances, or both. Specifically, an item can be judged as having come from a target list if the familiarities of target items are generally lower or higher than those of items from the other list. Similar familiarity-only models of list discrimination have been proposed to account for performance in experiments in which the strength of the items from two studied lists was expected to differ (Greene, 1999; Maddox & Estes, 1997).

Of particular interest is whether recollection is used when items from different sources have different levels of familiarity. It would seem that individuals should use recollection if distributions overlap greatly and if performance based on familiarity alone is expected to be low. However, this would equally apply for item recognition. If recollection is not used to discriminate studied items from nonstudied items that differ in strength, why should it be used to discriminate studied targets from studied lures that differ in strength?

Suppose, for example, that two study conditions, *A* and *B* serve to increment the familiarity of items to different levels. Specifically, the familiarity of *A* items after study is higher, on average, than the familiarity of *B* items. If the participants are later shown a mixed list of *A*, *B*, and new items and are instructed to endorse only those items from one source, a familiarity assessment should be able to support the discrimination of targets from lures.

If, for example, the instructions are to endorse only items from List *A* (i.e., strong *A* targets) and to reject all items from List *B* (i.e., weak *B* lures) and all new items, the decision can be made by simply accepting all items with familiarities that fall above a criterion (see Figure 1A). In this situation, discriminating between items from two lists is similar to discriminating between old and new items, because a familiarity value below the criterion is partially diagnostic of both new items and items from the lure source. If confidence ratings are also obtained, it can also be assumed that they are arranged along the familiarity axis, as is shown in Figure 1A, with highly confident endorsements on the high-familiarity end (e.g., "6") and highly confident rejections on the low-familiarity end (e.g., "1").

However, if participants are to endorse items from the less familiar source (i.e., weak *B* targets), they must discriminate target items not only from new items with lower familiarities, but also from old lures with higher familiarities (i.e., strong *A* lures). In this case, a unidimensional signal detection process with one response criterion is insufficient to produce old/new discrimination and list discrimination simultaneously.

One possible solution is to assume that individuals adopt two sets of response criteria, as is shown in Figure 1B. In
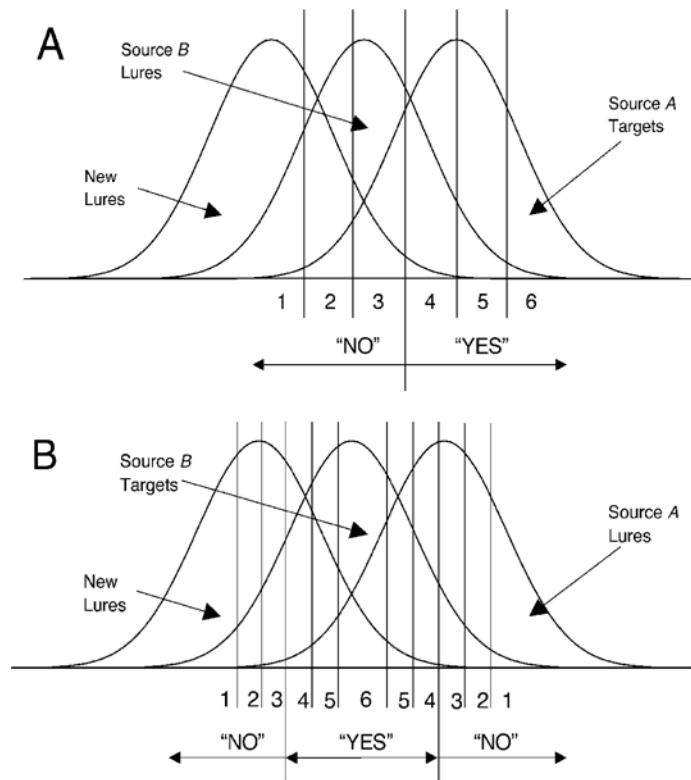
Figure 1. Hypothetical familiarity curves for strong studied items, weak studied items, and new items. Panels show (A) a test situation in which items from the stronger source are targets and participants use a single criterion to discriminate strong targets from weaker studied and nonstudied lures and (B) a test situation in which items from the weaker source are targets and participants use two criteria, a lower criterion that is used to discriminate studied targets from weaker nonstudied lures and an upper criterion that is used to discriminate weak studied targets from strong studied lures.

this model, a low criterion is chosen, below which items are rejected as new, and a high criterion is chosen, above which items are selected as having come from the non-target source. An endorsement is given to items with familiarities above an old/new criterion but below an *A/B* source criterion. Similar two-criterion strength models have been proposed to account for source discrimination tasks, such as reality monitoring (Hoffman, 1997; Marsh & Bower, 1993).[2] The two-criterion model has the implication that highly familiar items can be rejected if participants are given instructions only to endorse items from a less familiar source. Confidence ratings would not be arranged from low to high familiarity going left-to-right, as in Figure 1A, but rather as increasing and then decreasing, as in Figure 1B. Specifically, the most confident endorsement (i.e., "6") would be given not to all items above a high level of familiarity, but to items that fall within an intermediate familiarity range. Less confident endorsements are given to items both above and below the range, until the old/new and *A/B* criteria are reached, and then more and more confident rejections are given to items falling below the old/new criterion and above the *A/B* cri-

terion. Thus, each confidence point on the ROC represents the placement of two criteria along the familiarity axis.

Source Memory Discrimination Based on Recollection and Familiarity

The DPSD model can account for source discrimination, whether or not familiarity contributes to performance, by assuming that an item's source can also be recollected. Items can be judged as being from Source *A* if that information is recollected, even if the familiarity distribution of Source *A* items is no higher or lower than that of Source *B* items. The DPSD equations for predicting response probabilities for items from a target source, a lure source, and new items will be given below. Consider first a discrimination task in which one must endorse only items from a strong source. The probability of identifying a strong target item as a target (hit rate) is equal to the probability that it is recollected, R_s, plus the probability that it is judged to be familiar, F_s, in the absence of recollection:

$$P(\text{"target"} \mid \text{strong target}) = R_s + (1 - R_s)F_s. \quad (1)$$

The probability of a false alarm to an item from the weak source (i.e., the lure source) is equal to the probability that the lure is judged to be familiar, F_w, if the item is not recollected, R_w:

$$P(\text{"target"} \mid \text{weak lure}) = (1 - R_w)F_w. \qquad (2)$$

Importantly, Equation 2 assumes that recollection of a studied item is used to reject the item, whereas in Equation 1, recollection is used to accept the item. By contrast, items judged to be familiar are accepted in both equations. Note also that the model assumes that false alarms are always the result of incorrect familiarity assessments and that there is no provision for false recollection (e.g., recollecting that a weak source item came from a strong source). The probability of a false alarm to a new item is also equal to the probability that it is judged familiar, F_n, provided it is not recollected, R_n:

$$P(\text{"target"} \mid \text{new}) = (1 - R_n)F_n. \qquad (3)$$

Although in most cases participants cannot truly recollect that an item was not presented, there may be circumstances under which qualitative information may be used to reject new items (see Rotello et al., 2000, and Yonelinas, 1997, for discussions). For simplicity, R_n was assumed in the present experiment to be zero, leaving the false alarm rate for new items equal to F_n.

Familiarity is assumed to be well described by signal detection theory, so the probability that an item of any type will be judged familiar is equal to the area underneath the part of its familiarity distribution that exceeds the response criterion, $P(F > C_i)$. Familiarity distributions are assumed to be normal, and the mean of the new-item distribution is scaled to zero with unit variance. Thus, the criterion settings, as well as the means of target and lure distributions, are equal to their respective distances from zero in standard deviation units (i.e., z-scores) of the new distribution. The means of the strong and the weak distributions are expressed as d'_s and d'_w, respectively. Equations 1, 2, and 3 can be rewritten accordingly (i.e., as Equations 4, 5, and 6, respectively), so that the probability of judging any type of item to come from the target distribution is expressed as $\Phi[(d'_y - C_i)/S_y]$, the cumulative area of normal distribution y that exceeds a response criterion, C_i (one of k criteria used to construct the ROC, $C_1, C_2, \ldots, C_k$), given the mean, d'_y (0, d'_s, or d'_w), and standard deviation, S_y (1, S_s, or S_w), of the distribution:

$$P(\text{"target"} \mid \text{strong target}) =$$
$$R_s + (1 - R_s) * \Phi\left(\frac{d'_s - C_i}{S_s}\right), \quad (4)$$

$$P(\text{"target"} \mid \text{weak lure}) = (1 - R_w) * \Phi\left(\frac{d'_w - C_i}{S_w}\right), \quad (5)$$

$$P(\text{"target"} \mid \text{new}) = \Phi(-C_i). \qquad (6)$$

The familiarity terms for these equations are somewhat different for predicting responses when items from the target source are weak and those from the lure source are strong (see Figure 1B). When targets have intermediate strength relative to other items, endorsements (i.e., coming from the target distribution) are given to items with familiarity values above a lower criterion, C_i, but below an upper criterion, C_j. Thus, the probability of endorsing an item under these circumstances is the area of the familiarity distribution between C_i and C_j. Equations 4, 5, and 6 are modified accordingly, so that

$$P(\text{"target"} \mid \text{weak target}) =$$
$$R_w + (1 - R_w) * \left[\Phi\left(\frac{d'_w - C_j}{S_w}\right) - \Phi\left(\frac{d'_w - C_i}{S_w}\right) \right], \quad (7)$$

$$P(\text{"target"} \mid \text{strong lure}) =$$
$$(1 - R_s) * \left[\Phi\left(\frac{d'_s - C_j}{S_s}\right) - \Phi\left(\frac{d'_s - C_i}{S_s}\right) \right], \quad (8)$$

and

$$P(\text{"target"} \mid \text{new}) = \Phi(-C_j) - \Phi(-C_i). \qquad (9)$$

The equations represent a hybrid of both UVSD and DPSD models in that they contain all the parameters of both models. The UVSD and DPSD models make equivalent predictions if all the recollection terms are zero ($R_s = R_w = .00$) and if all the familiarity distributions have the same standard deviation ($S_s = S_w = 1.00$). In this case, the model reduces to a standard EVSD model. Under DPSD assumptions, the standard deviation terms are equal, but R_s and R_w are free to assume any value between zero and one. Free estimation of R for target items allows the ROC to intercept the ordinate between 0,0 and 0,1. Free estimation of R for lures allows the top end of the ROC to intercept the upper abscissa between 0,1 and 1,1. R was originally incorporated to account for the asymmetry commonly observed in ROCs (see Yonelinas, 1994), but if both R parameters are equal, the ROC predicted by the DPSD model is symmetrical. Under UVSD assumptions, recollection terms are fixed at zero, but the standard deviations of old-item familiarity distributions (S_s, S_w) are free to assume any value. Accordingly, the ROC will become "pulled" toward the axis of the distribution with the larger standard deviation. However, the function will always pass through 0,0 and 1,1, given that recollection terms are zero.

Proponents of both models have argued that the unique assumptions of the other are unnecessary. Statistically, however, the DPSD and the UVSD models cannot be directly compared, because they are not nested models. For example, when moving from the UVSD model to the DPSD model, two parameters must be fixed (S_s, S_w), and two parameters must be freed (R_s, R_w). However, because recollection and signal detection parameters are independent and serve different functions, there is no a priori reason why dual-process and unequal-variance assumptions cannot both be made by the same model. Indeed, an indirect way to compare two models statistically is to compare both to a less restricted hybrid model that contains all

the free parameters of the other two. The two models of interest can be derived from the hybrid model by fixing values of critical parameters. For example, a hybrid of the DPSD and the UVSD models would embody both dual-process assumptions (i.e., free parameters R_s, R_w) and unequal-variance assumptions (i.e., free parameters S_s, S_w); the UVSD model is derived by fixing recollection parameters to .00, and the DPSD is derived by fixing distribution variances to 1.00. Then, by comparing the UVSD and the DPSD models with the hybrid, it can be determined to what degree the parameter restrictions specific to each model are unacceptable and serve to worsen the fit. Specifically, a comparison of the hybrid with the UVSD model tests whether there is an improvement in fit by adding recollection if the variances are already allowed to be unequal. A comparison of the hybrid with the DPSD model tests whether there is an improvement in fit by allowing variances to be unequal when recollection already contributes to performance. Note that any hybrid model of this sort must contain more free parameters than either of the critical models and, as such, is less parsimonious than the others. Thus, fitting this type of model requires more than the usual number of ROC points.

In the present study, an attempt was made to determine the relative necessity of unequal-variance and dual-process assumptions in simultaneously accounting for recognition of source and occurrence when familiarity should be sufficient to make both types of discriminations. Specifically, we examined the degree to which participants would use recollection to discriminate target items from both studied lures and nonstudied items when the familiarity of target items differed from both types of lures. For all the participants, two study lists were presented with different encoding instructions so that items from the two lists differed in strength. However, at test, one group of participants was required to endorse only items from the stronger source, whereas the other group was required to endorse only items from the weaker source. Thus, the participants in the two conditions were expected to use different decision criteria, but the actual memory parameters for strong and weak lists were not expected to differ across conditions. DPSD and UVSD models were each fit to data from the two conditions simultaneously, and critical model parameters were constrained to be identical across conditions. Fits of the two models were examined to determine which one provided the best overall account of performance, and both models were compared with a hybrid model to determine the consequences of removing recollection or unequal-variance assumptions. If recollection is unnecessary for source recognition when familiarity can be used, the fit of the hybrid should not improve relative to that of the UVSD model. Conversely, the fit of the hybrid should improve relative to that of the UVSD model if recollection is used for source recognition even when familiarity is sufficient. If equal variances of familiarity distributions are sufficient for the familiarity process once recollection is ac-

counted for, the fit of the hybrid should not be an improvement over that of the DPSD model. Conversely, if unequal variances are a necessary feature of the familiarity process, the fit of the hybrid should be an improvement over that of the DPSD model.

METHOD

Participants

Thirty-six undergraduates from the University of California, Davis participated for course credit.

Materials

Two hundred forty facial sketches of famous people were selected from the Corel Draw Clip Art Portrait Dictionary. Faces were chosen that were judged likely to be familiar and recognizable to most college students. The faces chosen generally represented the areas of politics, business, movies and television, sports, and intellectualism. Faces were divided into three sets of 80 each. One set was presented in each of two study phases, and the faces in the remaining set were used as nonstudied lures in the recognition test. Across participants, there were six counterbalanced assignments of sets to conditions and four different presentation orders for each of the assignments. Thus, across participants, each face appeared equally often in all three phases. All the faces were presented to the participants on a 15-in. color computer monitor.

Procedure

All the participants received two study phases and a single test phase. In the first study phase, the participants viewed 80 faces and made rapid gender judgments about each of them (i.e., the weak list). Gender judgments were expected to constitute a shallow encoding task and lead to relatively weak encoding, because a correct response required only the processing of structural features. The z and /keys were used to respond *male* and *female*, respectively. The participants were told to take enough time to correctly classify each face as male or female but to use no more than the time required to respond correctly. They were also instructed to respond carefully, because some male faces had long hair and some female faces had short hair.

In the second study phase, the participants viewed a second set of faces and judged the area in which they felt the person was most famous (i.e., the strong list). The participants were instructed to classify each face according to which of seven categories that person best represented. Response categories included business figures, leaders (political, military, spouse of leader), performers (actor, singer, model), media personalities (TV host, newscaster), sports figures (athlete, coach, sportscaster), and intellectuals (scientist, author, artist). The participants placed each face in one of the seven categories by pressing a corresponding number on the keyboard. They were instructed to classify each face only if they actually knew who the person was, and not to simply guess on the basis of some other information (e.g., the person's attire). The participants were told that they did not need to know an individual's name but that they needed only to recognize that person as fitting into one of the categories. If, however, they did not know the person, they pressed the "8" key. A table with the assignment of categories to numbers was always presented along with a face. As much time was allowed as was necessary for this decision phase. Following each classification response, the participants were told whether or not they were correct and were given a brief description on the computer screen of why the person was well known. On the basis of this information, they were again asked whether they knew the correct classification. The participants were free to modify their first responses at this point before continuing to the next face. Fame judgments and the subsequent elaborate feedback procedure were

Table 1
Response Frequencies for Six Levels of Confidence
and Three Test Lists in Two Conditions

Condition	Item	6	5	4	3	2	1	Total
Weak targets	strong	91	126	119	149	206	749	1,440
	weak	252	274	196	208	206	304	1,440
	new	49	93	144	253	261	640	1,440
Strong targets	strong	613	298	168	127	118	116	1,440
	weak	117	127	131	243	365	457	1,440
	new	47	61	98	266	360	608	1,440

Column header spanning 6–1: Response

expected to constitute a relatively deep encoding task that should lead to relatively strong encoding, because the participants' involvement in the task was greater than in the first encoding phase.

Following the two study phases, all the participants were given a recognition memory test. The test consisted of all 160 previously studied faces, along with 80 new faces. Half of the participants were instructed to endorse only the faces for which they had previously made sex judgments (weak target condition) and to reject both nonstudied faces and faces for which they had made fame judgments (i.e., both strong and new lures). The other half of the participants were instructed to endorse only the faces for which they had made fame judgments (strong target condition) and to reject both new faces and faces for which they had made sex judgments (i.e., both weak and new lures). Thus, the conditions differed according to which list the participants were instructed to endorse. Each face was rated according to how confident the participant was that it had appeared in the target phase (sex or fame), with ratings ranging from 1 (*certain that it did not appear in the target phase*) to 6 (*certain that it did appear in the target phase*). Responses of 1 or 6 indicated that the participant had complete confidence, 2 or 5 indicated moderate confidence, and 3 or 4 indicated low confidence. The participants were further instructed to spread out their responses so that they used the entire range.

ANALYSIS

Each participant contributed 240 responses distributed among 18 bins. Table 1 shows the total frequency count summed across participants for each of the 18 bins in each of the two conditions. Each frequency bin is associated with one of the six confidence judgments made for faces in one of the three lists (strong, weak, and new). Response frequencies were converted to probabilities for five different criterion settings by cumulating across confidence levels for each list. Performance at the most conservative criterion setting was estimated as the proportion of faces receiving only the most confident endorsements (i.e., only those faces receiving a 6). Probabilities for the second most conservative setting were computed by adding the proportion of faces receiving a response of 5 to the proportion of faces receiving a response of 6. Successively less conservative settings were computed in this manner, so that 15 endorsement probabilities were obtained from each participant that corresponded to five criterion settings for each of the three types of faces.

Fitting the Models

ROCs are often assessed by examining the fits of regression models that predict hit rates from false alarm rates. This procedure is problematic in that it yields a squared-error term that is computed only from variability of the hit rate and assumes that false alarm rates are perfectly known. That is, because the false alarm rate is treated as a predictor variable, the best-fitting ROC function minimizes only the squared deviations from the hit rate. Both hit and false alarm rates contain error variance, and thus, we used a minimization function that reflected variability of both hit and false alarm rates. The model fitting was performed using the solver function in Microsoft Excel 7.0 (see Dodson, Prinzmetal, & Shimamura, 1998; Yonelinas, 1994). The solver employs an efficient search algorithm that can be used to minimize the value of a specified cell in an Excel spreadsheet by changing the values of other specified cells. The value in the to-be-minimized cell can be defined as a function of the values in to-be-changed cells. In this manner, the solver was used to minimize a total *SSE* term that was the sum the *SSE*s of all the axes. The function to be minimized was defined as

$$SSE = \sum_i^3 \sum_j^5 \sum_k^2 \left(P_{ijk} - \hat{P}_{ijk}\right)^2, \qquad (10)$$

where the squared deviations between observed (P_{ijk}) and predicted ($\hat{P}_{ijk}$) response probabilities are summed over $i = 3$ item types (strong, weak, and new items), $j = 5$ criterion settings (cumulated confidence levels), and $k = 2$ experimental conditions. Thus, the models were fit to the observed data from the strong- and the weak-target conditions simultaneously by minimizing an overall *SSE* term that was summed not only across confidence levels, but also across the three types of items and the two conditions. Because each participant contributed 5 response probabilities for each type of item, there were 15 averaged probabilities from each of the two conditions, for a total of 30 data points. Thus, the overall *SSE* term was a sum of 30 squared errors between observed and model-generated probabilities. This procedure may be best conceptualized as two simultaneous three-dimensional regressions for which the *SSE* term reflects deviations of five predicted points from five observed points on each of the *x*-, *y*-, and *z*-axes (target source, lure source, and new items, respectively) in each condition.

The solver was used to find the best-fitting values of the signal detection criterion settings (C_i), as well as critical model parameters (R_s, R_w, d'_s, d'_w, S_s, S_w). Model-generated response probabilities for strong, weak, and new items for each level of confidence were computed using the same criterion values. However, a different criterion setting was estimated for each level of confidence within a given condition, and different settings were estimated for the two conditions. Thus, each criterion setting was estimated to generate a predicted response probability (i.e., a confidence point) on all three axes of one of the two conditions. Note, however, that equality constraints were imposed on the critical model parameters across the two conditions. For example, regardless of whether test instructions requested the endorsement of strong fame items or weak sex items, it was assumed that all the weak sex items would have the same average familiarity across test

instructions and that the stronger fame items would have the same average familiarity across test instructions. Equality constraints were made on the critical parameters of both models.

The DPSD and UVSD models both contained 19 free parameters. Of these, 17 were shared by the two models. Two parameters were values of d' for strong and weak items. The other 15 free parameters shared by all the models represented the criterion settings for the two conditions. Criterion settings were assumed to be under participant control and were allowed to vary across the two conditions. Whereas only 1 setting was assumed to be necessary for each confidence point in the strong-target condition (see Equations 4–6), 2 settings were assumed to be necessary for each confidence point in the weak-target condition (see Equations 7–9). Thus, 5 settings were estimated from the strong-target condition, 1 for each level of confidence, and 10 settings were estimated from the weak-target condition, 2 for each level of confidence. To reiterate, faces given fame judgments were expected to lead to higher strength than were faces given sex judgments, and when instructed to endorse only those faces that had been in the fame list (i.e., strong targets), the participants needed only to accept the most familiar items as having been studied. Thus, the actual decision process and criterion settings in this condition were considered to be similar to those of a simple old–new discrimination. However, if the participants were instructed only to endorse items that had received a sex judgment (i.e., weak targets), one upper criterion and one lower criterion were assumed to be necessary for each level of confidence, because the participants were required to accept only items of intermediate familiarity.

For the DPSD model, two additional parameters were used to represent the contribution of recollection to strong and weak items (R_s and R_w, respectively). Both R values were constrained to assume values between zero and one. For the UVSD model, two additional parameters were freed for the standard deviations of strong and weak familiarity distributions (S_s and S_w, respectively). Equality constraints across conditions were imposed on R_s and R_w in the DPSD model and on S_s and S_w in the UVSD model. Both S parameters were constrained to assume a value of 1.00 or greater.[3] For the hybrid model, with which the DPSD and the UVSD were compared, R_s, R_w, S_s, and S_w were all estimated under the same constraints.

The fits of the models were evaluated by examining the R^2 value for the fitted solution.[4] The R^2 was computed using the overall SSE term and the sum of the six total sums of squared deviations (SST_{ik}) for the three item types in the two conditions. This SST term was defined as

$$SSE = \sum_i^3 \sum_j^5 \sum_k^2 \left(P_{ijk} - \overline{P}_{ik} \right)^2, \qquad (11)$$

where squared deviations of observed points (P_{ijk}) from an item/condition mean ($\overline{P}_{ik}$) were summed across $i = 3$ item types, $j = 5$ confidence points, and $k = 2$ experimental conditions. Note that in this equation, deviations of each point from its corresponding item/condition mean

are used, rather than deviations from a grand mean. That is, each model fit is assessed by the degree to which it explains the sum of six separate variance estimates (x-, y-, and z-axes in the two conditions), rather than the total variability among the 30 points. This procedure was deemed appropriate given the multidimensional nature of the data (i.e., a separate variance measure was obtained for each dimension). Using deviations from a grand mean gave larger R^2 estimates, but significance tests for model comparisons were nearly identical for the two procedures.

As was discussed earlier, a direct comparison of R^2 values for the DPSD and the UVSD models could not be statistically assessed, because the models are not nested. Thus, nested comparisons were made between each of these models and the hybrid model that contained both R and S parameters. A more complete description of this analysis is given later. Also, because the observed data points represented proportions of response frequencies summed across participants, 95% confidence intervals about the observed data were examined as well. For each model, the number of predicted points was counted that fell outside a 95% confidence interval about the observed points. Regardless of the R^2, any model fit should be rejected if there is a significantly greater number of predicted points falling outside the confidence intervals than would be expected given a true 95% success rate. A .05 alpha level was adopted for all significance tests.

RESULTS

The average ROCs are presented in Figures 2 and 3, along with the model-generated data points. Hit rates appear on the y-axes, and false alarm rates appear on the x-axes. Three plots are shown for each condition, one for each possible pairwise relationship between item types. For the weak-target condition, weak sex targets are plotted against strong fame lures (panel A) and against new items (panel B). Strong fame lures are plotted against new items in panel C. For the strong-target condition, strong fame targets are plotted against weak sex lures (panel D) and against new items (panel E). Weak sex lures are plotted against new items in panel F. Ninety-five percent confidence intervals are given around each observed point. Note that because the R^2 of these fits reflect variability of all three item types, confidence intervals were computed for all item types and are shown for variables on both axes in each ROC. Thus, the intervals in Figures 2 and 3 each represent one 2-dimensional side (i.e., a rectangle) of a confidence "cube." Model-generated points, rather than continuous functions, are shown in the figures to illustrate their locations within each confidence rectangle. Note that because the fitting occurred for three dimensions (sex, fame, and new items), it is not enough that the model-generated function pass through the two-dimensional confidence rectangle, as is depicted in the individual ROCs. Rather, the model-generated points must lie within the three-dimensional confidence cube to provide an adequate fit. This is because each model-generated point repre-
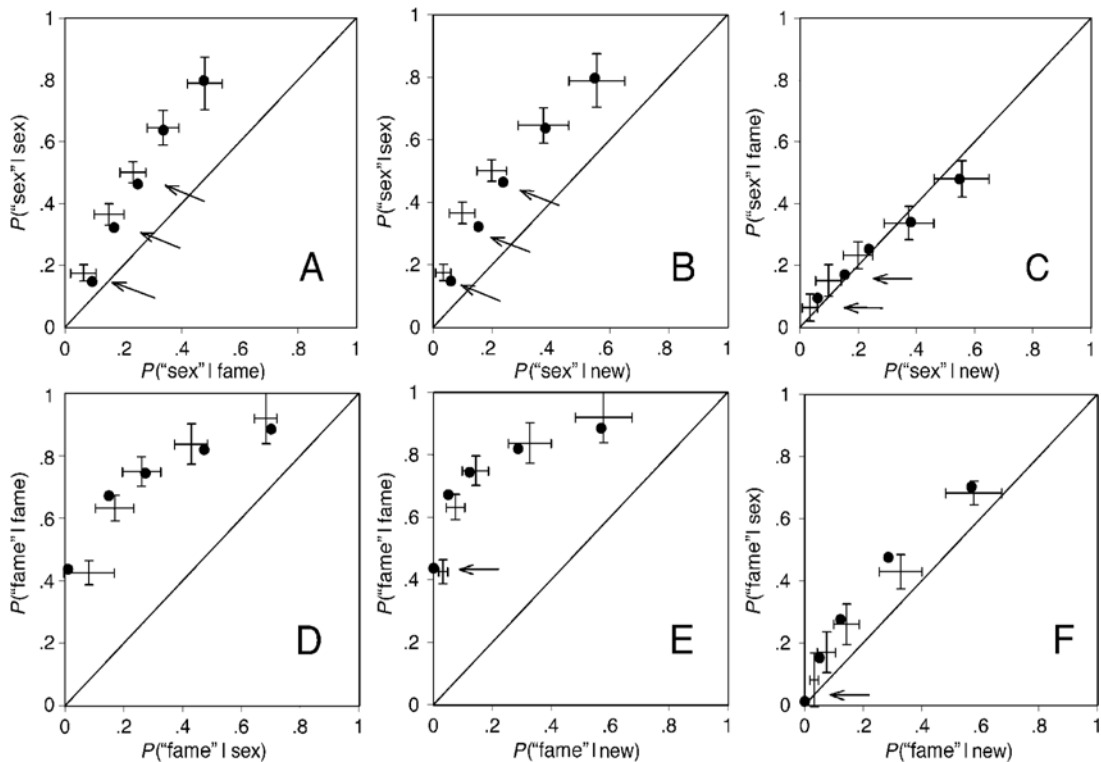
Figure 2. Observed and model-generated data points for the unequal-varaince signal detection model displayed as points on an ROC. Error bars about each point represent 95% confidence intervals. Arrows indicate ROC points where model predictions fall outside the confidence intervals for at least one of the two axes. Panels A, B, and C show ROCs from the weak-target/strong-lure condition for (A) list discrimination with weak sex targets plotted against strong fame lures, (B) item discrimination with weak sex targets plotted against new lures, and (C) strong fame lures plotted against new lures. Panels D, E, and F show ROCs from the strong-target/weak-lure condition for (D) list discrimination with fame targets plotted against weak sex, (E) item discrimination with strong fame targets plotted against new lures, and (F) weak sex lures plotted against new lures.

sents the closest part of the function to the observed point in three dimensions. A proper representation of such a fit would require a three-dimensional scatterplot with a model-generated function passing through confidence cubes around each point. A function that passed through a rectangle in two dimensions (e.g., strong targets and weak lures) might still pass outside the confidence interval about a value on the third dimension (e.g., new lures) and would miss the cube.

The Unequal-Variance Signal Detection Model

Data points generated by the UVSD model are shown in Figure 2, along with observed points and confidence intervals. The ROCs shown in panels A and D represent source discrimination (i.e., target faces plotted against studied lures). Panel A shows the source ROC for the participants who received instructions to endorse only items from the sex list (i.e., the weak-target condition), and panel D shows the source ROC for the participants who received instructions to endorse only the items from the fame list (i.e., the strong-target condition). The ROCs shown in panels B and E represent the ability to discriminate studied targets from nonstudied lures. Panel B shows the target/nonstudied ROC for the weak-target condition

(sex targets vs. new lures), and panel E shows the target/nonstudied ROC for the strong-target condition (fame targets vs. new lures). Finally, ROCs comparing false alarms to studied lures versus nonstudied lures are shown in panels C and F. Panel C shows the ROC for the weak-target condition (fame lures vs. new lures), and panel F shows the ROC for the strong-target condition (sex lures vs. new items).

The UVSD model accounted for 97.6% of the total observed variance in the average ROCs and for between 96.5% and 98.4% of the variance of any of the three lists for either of the two conditions. The best-fitting UVSD parameter estimates are shown in Table 2. Under the assumptions of the UVSD model, making fame judgments about faces during study led to greater discrimination of studied items from new lures ($d'_s = 2.70$) than did making sex judgments ($d'_w = 0.74$). The ability to discriminate between items in the weak list and items in the strong list according to the UVSD model is represented by a substantial d' as well ($d'_s - d'_w = 1.96$). Also, the standard deviation of the strong familiarity distribution ($S_s = 2.07$) showed an increase relative to that of the weak distribution ($S_w = 1.00$). These estimates are substantively consistent with the general finding that the ratio of nonstudied to studied
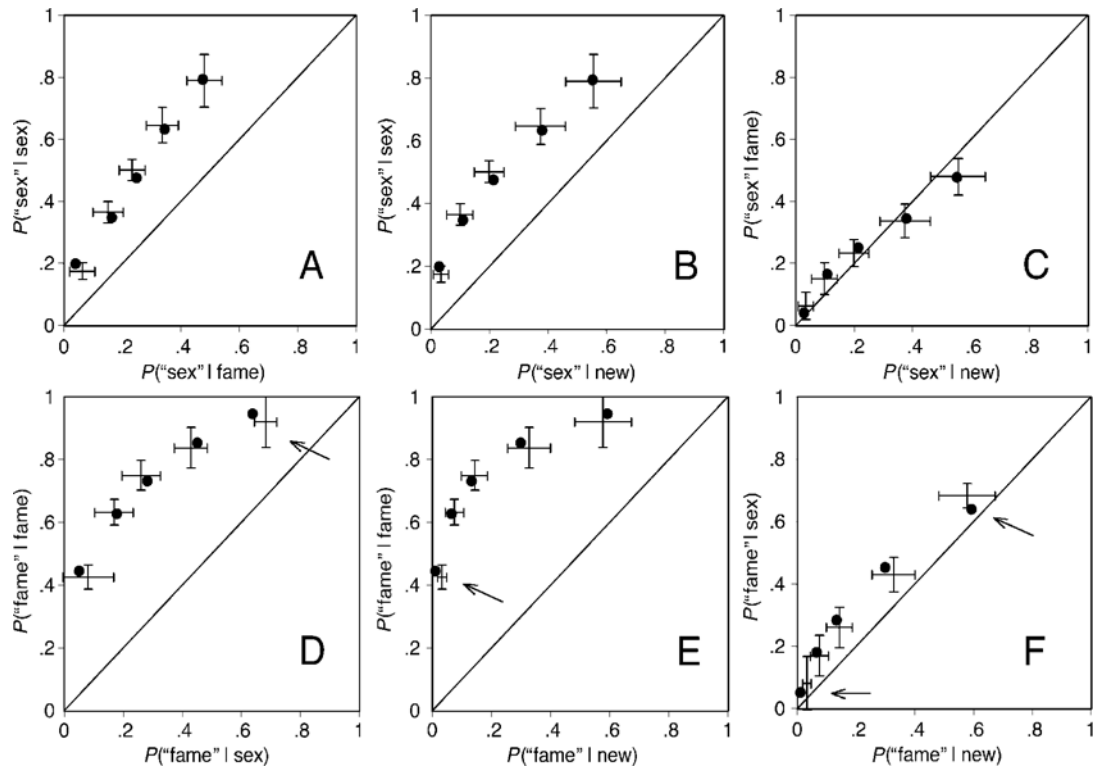
Figure 3. Observed and model-generated data points for the dual-process signal detection model displayed as points on an ROC. Error bars about each point represent 95% confidence intervals. Arrows indicate ROC points where model predictions fall outside the confidence intervals for at least one of the two axes. Panels A, B, and C show ROCs from the weak-target/strong-lure condition for (A) list discrimination with weak sex targets plotted against strong fame lures, (B) item discrimination with weak sex targets plotted against new lures, and (C) strong fame lures plotted against new lures. Panels D, E, and F show ROCs from the strong-target/weak-lure condition for (D) list discrimination with strong fame targets plotted against weak sex lures, (E) item discrimination with strong fame targets plotted against new lures, and (F) weak sex lures plotted against new lures.

standard deviations tends to decrease with increasing accuracy ($d'_w = 0.74$, $S_n/S_w = 1.00$ vs. $d'_s = 2.70$, $S_n/S_s = 0.48$; see Glanzer et al., 1999).

Figure 2 also shows model-generated points that represent the best-fit approximations to the observed data points. Arrows indicate ROC data points where the model predictions fall outside the confidence intervals for at least one of the two plotted axes. Six of the 30 predicted values fall outside a 95% confidence interval of their respective observed values (binomial $p = .0033$). In the weak-target condition, the first three points fall below the intervals for the weak sex targets, and the first two fall below the intervals for the new lures (see panels A, B, and C). In the strong condition, the first point falls below the interval for new lures (see panels E and F). Thus, 80% (24 of 30) of the UVSD model-generated data points fall within the 95% confidence intervals around the observed points.

Dual-Process Signal Detection Model

Data points generated by the DPSD model are shown in Figure 3, along with observed points and confidence intervals. The panels in Figure 3 reflect the same plots as

those shown in Figure 2, but with best-fitting approximations generated by the DPSD model rather than by the UVSD model. The DPSD model accounted for 99.0% of the total observed variance and between 98.3% and 99.7% of the variance of any of the three item types for either

Table 2
Least-Squares Parameter Estimates
and Fit Indices of DPSD and UVSD Models

Model	Fitted Parameter Values						Fit Indices		
	R_s	R_w	d'_s	d'_w	S_s	S_w	R^2	df	%CI
UVSD	.00*	.00*	2.70	0.74	2.07	1.00	.976	11	80
DPSD	.30	.16	1.58	0.87	1.00*	1.00*	.990	11	93
Hybrid	.00	.14	2.17	0.85	1.40	1.00	.993	9	97

Note—DPSD, dual-process signal detection model; UVSD, unequal-variance signal detection model; Hybrid, hybrid dual-process unequal-variance signal detection hybrid model; R_s, probability of recollecting an item from the strong list; R_w, probability of recollecting an item from the weak list; d'_s, signal detection discriminability of strong items from new items; d'_w, signal detection discriminability of weak items from new items; S_s, standard deviation of strong-item familiarity; S_w, standard deviation of weak-item familiarity. Range restrictions were imposed on R_s and R_w ($0.00 \geq R \geq 1.00$) and on S_s and S_w ($S \geq 1.00$). %CI, percentage of predicted data points that fall within a 95% confidence interval. *Fixed parameter value.

of the two conditions. The best-fitting DPSD parameter estimates are shown in Table 2. Fame judgments about faces during study led to both a higher probability of recollection ($R_s = .30$), and a greater familiarity discrimination of studied items from new items ($d'_s = 1.58$) than did making sex judgments ($R_w = .16$; $d'_w = 0.87$). As was predicted, the participants used familiarity to discriminate between strong and weak items ($d'_s - d'_w = 0.71$). These estimates are substantively consistent with the general finding that levels-of-processing manipulations affect both recollection and familiarity (see Yonelinas, 2001).

To determine whether it was reasonable to assume that R_s and R_w were each equivalent across conditions, the equality constraints were removed for these parameters, and the model was refit. Two degrees of freedom were lost by allowing R_s and R_w to vary across conditions. The modified model explained 99.3% of the variance and did not represent an improvement over the original model [$F(2,9) = 1.72$, $MS_e = 0.000834$, $p = .234$]. As compared with the strong-target condition, R_s was larger for the weak-target condition (.47 vs. .19), and R_w was slightly larger for the weak-target condition (.19 vs. .15). Thus, for the present experiment, there may have been a tendency for the participants to use recollection more in the weak-target condition, but the fit of the model was not significantly compromised by assuming that recollection was used equivalently in the two conditions.

The model-generated points in Figure 3 represent the closest approximation of the DPSD model to the observed data points. Of 30 estimated data points, 2 predicted values fall outside a 95% confidence interval of their respective observed value (binomial $p = .45$). In the weak-target condition, the last point falls below the interval for weak targets (see panels D and F), and the first point falls below the interval for the new lures (see panels E and F). Thus, 93% (28 of 30) of the model-generated data points fall within 95% confidence intervals about the observed points.

A comparison of the two models suggests that the DPSD model provided a better account for the ROC data than did the UVSD model. For the DPSD model, only 2 out of 30 points fell outside the 95% confidence intervals, whereas for the UVSD model, 6 out of 30 points fell outside the confidence intervals.

Dual-Process Unequal-Variance Signal Detection Hybrid

A hybrid model was specified that contained all the free parameters of DPSD and UVSD models. The hybrid model is simply a dual-process model without the equal-variance assumption for the familiarity distributions. Thus, the hybrid model contains two more free parameters than do the other two models. Comparing this model with the other two models allows two assumptions to be tested. Comparison with the UVSD model tests the assumption that given unequal variances, recollection is not necessary to account for performance. Comparison with the DPSD model tests the assumption that given the contribution of recollection, familiarity distributions have equal variance.

The hybrid model accounted for 99.3% of the total variance across both conditions. The fit was significantly better than that of the UVSD model [$F(2,9) = 10.98$, $MS_e = 0.000847$, $p = .0039$], indicating that even with a provision for unequal variances, the fit was improved when recollection was added to the model. By contrast, the fit was not significantly better than that of the DPSD model [$F(2,9) = 1.69$, $MS_e = 0.000847$, $p = .23$], indicating that no significant improvement in fit was gained over that of the dual-process model by allowing unequal variances for familiarity distributions. Thus, a comparison of both models with the hybrid reveals that the equal-variance restriction of the DPSD was reasonable and did not compromise the fit of that model. Conversely, the zero-recollection assumption of the UVSD was unreasonable and did compromise the fit of the model.

The best-fitting parameters of the hybrid model are presented in Table 2. Taken at face value, the parameters suggest that recollection occurred only for weak items ($R_s = .00$, $R_w = .14$) and that an increase in variance occurred only for strong items ($S_s = 1.40$, $S_w = 1.00$). This finding is substantively inconsistent with a body of literature on the effect of level of processing on estimates of recollection and familiarity (Yonelinas, 2001). On one hand, this appears to indicate that the improvement over the UVSD model was exclusively the result of adding R_w. However, because there was no improvement of the hybrid over the DPSD model, parameter differences between the hybrid and the DPSD models (.16 vs. .14 for R_w, .00 vs. .30 for R_s, 1.40 vs. 1.00 for S_s) should be interpreted with caution. Although a hybrid account of recognition ROCs certainly cannot be rejected on the basis of the present experiment, it did not appear to represent a substantive improvement over either the UVSD or the DPSD models and was not a statistical improvement over the DPSD model. It was, however, a statistical improvement over the UVSD model.

DISCUSSION

In the present study, unequal-variance and dual-process signal detection models were fit to recognition ROCs in conditions in which familiarity should have been useful for recognition of both source and occurrence. The participants discriminated either strong targets from weak studied lures and new lures or weak targets from strong studied lures and new lures. A simple familiarity process was modeled as a UVSD model and did not provide an adequate fit to the ROCs of the two conditions simultaneously. A DPSD model in which familiarity was augmented by recollection did provide an adequate fit of ROCs of the two conditions simultaneously. When both unequal variances and recollection were assumed for the same model, there was no improvement over the dual-process model, but there was an improvement over the unequal-variance model. The results suggest that recollection contributed to recognition in the present study even though above-chance performance could have been achieved by using only familiarity and that equal variances of familiarity

distributions were sufficient for the familiarity process once recollection was accounted for. The results further suggest that people do not limit their recognition decisions to familiarity assessments when familiarity information is useful but also make use of recollection when it is available to them.

The results are inconsistent with the view that recollection contributes to performance only when familiarity information is relatively ambiguous. Although it appears that a threshold process such as recollection is necessary to account for linear ROCs in associative and source recognition tasks, it has been argued that such findings are unusual and that the contribution of a threshold process is not a general characteristic of either item recognition (Glanzer et al., 1999) or source recognition (Qin et al., 2001; Slotnick et al., 2000). Familiarity-only signal detection accounts of recognition ROCs assume unequal variances of familiarity distributions in order to account for the asymmetry of the ROCs but must also assume that individuals employ a *switching* strategy to account for linear ROCs. That is, a familiarity-only view assumes that whenever possible, participants make a familiarity assessment only on the basis of a continuous signal detection process. However, when familiarity cannot be used to make recognition decisions (e.g., when discriminating between items of two equally familiar sources), participants switch to a threshold-based recollection process. The present data contradict this type of switching view, in that both processes contributed to performance in the same task.

Note that other process-switching strategies may be consistent with the present data. For example, the dual-process account presented here can be viewed as a trial-by-trial switching model whereby participants either recollect an item on a given trial or, failing that, resort to a familiarity assessment. Thus, recollected information may contribute to a decision on one trial, whereas a decision on another trial may be based on a familiarity assessment.

In the present paper, we asked whether recollection was used, given that it was possible for the participants to make their decisions solely on the basis of familiarity. We considered this to be the reasonable approach, given that most models of recognition assume that familiarity is an important component of recognition performance. However, one could easily have asked the opposite question: Given that people use a recollection process, is there evidence that they also use a familiarity process? If it is assumed that recollection is a threshold process, nonlinear ROCs can be taken as evidence for a familiarity process. Figures 2 and 3 show that the functions are clearly curvilinear, which suggests that familiarity was used and that a simple threshold process would not be sufficient.

In the present study, a simple form of familiarity-only recognition was tested in the context of a complex recognition task. The following assumptions were made about familiarity: (1) Familiarity is a unidimensional signal detection process, and (2) participants are able to control response criteria only, and not the actual computation of familiarity. The first assumption distinguishes the present model of familiarity from multidimensional signal detection (MSD) models that propose that items, when studied, can become more familiar along more than one dimension (see Macmillan & Creelman, 1991). The second assumption about familiarity in the present model distinguishes it from some global-matching models of recognition, in which context parameters can be used to weight familiarity distributions according to test instructions (e.g., the matrix model, Humphreys, Bain, & Pike, 1989; SAM, Gillund & Shiffrin, 1984; see Clark, 1999, and Ratcliff et al., 1995). Both alternatives are discussed below.

Multidimensional Signal Detection Theory

The source recognition tasks presented here can be viewed in light of an MSD model of source recognition. For example, the present data could be accounted for by assuming that the two source lists are represented by specific strength axes and that studying an item serves to increase its strength along both axes to some degree. As such, all test items would have both a *sex list* strength value and a *fame list* strength value. Studying an item in the fame list would increase its fame strength to a greater extent than its sex strength, and vice versa for items in the sex list. It is assumed that to make a recognition decision, participants rotate a single decision axis to an angle so that the strength distributions, when projected onto the axis, are discriminable (i.e., the distance between target and lure distributions is maximized). Kinchla (1994) and Batchelder, Riefer, and Hu (1994) have summarized the advantages and disadvantages of MSD modeling, using MSD theory for source-monitoring data, and Banks and colleagues (Banks, 2000; Banks, Chen, & Prull, 1999) have applied MSD to the analyses of source recognition and to inclusion and exclusion performance in a process dissociation paradigm.

Although the proposals are interesting, MSD models run into the same problem as unidimensional signal detection models, in that they cannot account for linear ROCs. As such, it must be assumed that linear ROCs occur when participants abandon a signal detection process in favor of a threshold process for performance. Linear ROCs could be produced by introducing nonnormal distributions, but it is unclear what effect this modification would have on other aspects of the model (e.g., rotation of the decision axis).

The multidimensional memory representation assumed in MSD theory may, in some cases, be consistent with the memory representation assumed by the dual-process model. For example, consider a two-dimensional memory space in which the dimensions represent the strengths of two sources (e.g., fame strength vs. new, sex strength vs. new). If the space is rotated 45°, the dimensions will more likely resemble item strength (old vs. new) and source strength (fame vs. sex). The latter solution is more compatible with multinomial and dual-process models of source recognition than is the former, but both will provide equivalent statistical fits to the data, provided that

the item distributions do not change positions relative to one another. In fact, the dual-process model of Yonelinas (1994) is equivalent to a multidimensional model in which item distributions are Gaussian along a familiarity dimension and rectilinear along a recollection dimension. A decision is first made on the recollection dimension by determining whether an item falls above a region of overlap between old and new items (i.e., whether it exceeds a threshold). If not, the item's strength is independently assessed by an evaluation of its familiarity relative to a response criterion on the familiarity dimension.

In summary, MSD theory represents an innovative and powerful alternative to the present dual-process model, in that performance on a variety of source recognition tasks can be explained by the rotation of a single decision axis through a multidimensional memory space. However, the MSD models do not account for linear ROCs. Moreover, without some a priori consideration of the nature of the memory dimensions, it is difficult to determine just how substantively different an MSD model solution is from that of a dual-process or multinomial model solution. Much work needs to be done to determine whether MSD models of recognition are both practical and psychologically meaningful.

Global-Matching Models of Recognition

The second assumption about familiarity in the present model distinguishes it from global-matching models of recognition (e.g., SAM; Gillund & Shiffrin, 1984). In global-matching models of memory, the familiarity or strength is a single value that is computed as the match between the test item and all the items in memory. Both SAM and the matrix model (Humphreys et al., 1989) can account for certain types of source recognition (e.g., exclusion in the process dissociation procedure) by assuming that participants can "focus" retrieval on certain items and, thus, that the actual computation of a familiarity value is under partial control of the individual (Clark, 1999; Ratcliff et al., 1995). The use of context in the computation of familiarity is necessary in a single-process model because context can be retrieved and used to make recognition decisions (Humphreys et al., 1989; Ratcliff et al., 1995). This is accomplished in global-matching models by the addition of context parameters to *weight* items according to the retrieval instructions. If the test instructions require endorsing one set of studied items and rejecting another, items from the two sets can be given different context weights by the individual in an attempt to discriminate between them. By contrast, dual-process theories generally assume that the intentional use of context affects recognition decisions primarily through the recollection component and that the contribution of context to the computation of familiarity cannot be controlled on the basis of the test instructions alone. This is not to say that context cannot affect familiarity in a dual-process model. Rather, the assumption of dual-process models is that any intentional focusing of retrieval on items of a specific context is reflected in recollection, not in familiarity.

It is, of course, quite plausible that familiarity is sensitive to the specific goals and intentions of the participants. However, it may be more useful to first rule out the role of recollection in these effects, given previous research showing that familiarity behaves more like an automatic process than like a controlled process (e.g., Jacoby, 1991; Jacoby, Woloshyn, & Kelley, 1989; Rajaram, 1993). Future studies will be useful in testing this notion.

Although the assumption that familiarity is context specific may turn out to be true, the global-matching models are not consistent with recognition memory ROC data. First, some global models predict that recognition ROCs will be symmetrical, whereas others predict that they will always become less symmetrical as performance increases, and both of these predictions have been falsified (see Ratcliff et al., 1992; Yonelinas, 1994). Second, they produce Gaussian memory strength distributions, and thus they are inconsistent with studies demonstrating linear ROCs (e.g., Rotello et al., 2000; Yonelinas, 1997, 1999a). Whether it is possible to modify these models to bring them in line with existing ROC data is not yet clear.

REFERENCES

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Erlbaum.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In R. C. Atkinson, R. D. Luce, D. H. Krantz, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I. Learning, memory and thinking* (pp. 243-293). San Francisco: Freeman.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin*, **74**, 81-99.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, **11**, 267-273.

Banks, W. P., Chen, Y., & Prull, M. W. (1999). Memory and awareness: Is memory information stratified into conscious and unconscious components? In B. H. Challis & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 129-172). Amsterdam: Benjamins.

Batchelder, W. H., Riefer, D. M., & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, **101**, 172-176.

Clark, S. E. (1999). Recalling to recognize and recognizing to recall. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson–Shiffrin model* (pp. 215-243). Mahwah, NJ: Erlbaum.

Dodson, C. S., Prinzmetal, W., & Shimamura, A. P. (1998). Using Excel to estimate parameters from observed data: An example from source memory data. *Behavior Research Methods, Instruments, & Computers*, **30**, 517-526.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, **24**, 523-533.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, **16**, 309-313.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

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

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Greene, R. L. (1999). The role of familiarity in recognition. *Psychonomic Bulletin & Review*, **6**, 309-312.

HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

HIRSHMAN, E., & HENZLER, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, **9**, 61-65.

HIRSHMAN, E., & MASTER, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. *Memory & Cognition*, **25**, 345-351.

HOFFMAN, H. G. (1997). Role of memory strength in reality monitoring decisions: Evidence from source attribution biases. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 371-383.

HUMPHREYS, M. S., BAIN, J. D., & PIKE, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, **96**, 208-233.

INOUE, C., & BELLEZZA, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, **26**, 299-308.

JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.

JACOBY, L. L., WOLOSHYN, V., & KELLEY, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, **118**, 115-125.

KELLY, R., & WIXTED, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 701-722.

KINCHLA, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, **101**, 166-171.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

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

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

MANDLER, G. (1991). Your face looks familiar but I can't remember your name: A review of dual process theory. In W. E. Hockley (Ed.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 207-225). Hillsdale, NJ: Erlbaum.

MARSH, R. L., & BOWER, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 673-688.

MURDOCK, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

OGILVIE, J. C., & CREELMAN, C. D. (1968). Maximum likelihood estimation of receiver-operating characteristic curve parameters. *Journal of Mathematical Psychology*, **5**, 377-391.

QIN, J., RAYE, C. L., JOHNSON, M. K., & MITCHELL, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 1110-1115.

RAJARAM, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*, **21**, 89-102.

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, & Cognition*, **20**, 763-785.

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

RATCLIFF, R., VAN ZANDT, T., & MCKOON, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General*, **124**, 352-374.

ROTELLO, C. M., MACMILLAN, N. A., & VAN TASSEL, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language*, **43**, 67-88.

SLOTNICK, S. D., KLEIN, S. A., DODSON, C. S., & SHIMAMURA, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1499-1517.

SWETS, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, **99**, 100-117.

TULVING, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.

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

YONELINAS, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, **25**, 747-763.

YONELINAS, A. P. (1999a). 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, & Cognition*, **25**, 1415-1434.

YONELINAS, A. P. (1999b). Recognition memory ROCs and the dual-process signal-detection model: Comment on Glanzer, Kim, Hilford, and Adams (1999). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 514-521.

YONELINAS, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, **130**, 361-379.

YONELINAS, A. P., DOBBINS, I., SZYMANSKI, M. D., DHALIWAL, H. S., & KING, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness & Cognition*, **5**, 418-441.

YONELINAS, A. P., KROLL, N. E. A., DOBBINS, I. G., & SOLTANI, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review*, **6**, 654-661.

NOTES

1. Technically, threshold models predict linear ROCs for confidence ratings only when all the recollected items are given the highest confidence rating. Thus, the dual-process model assumes that all recollected items will lead to a highly confident endorsement or rejection. The high-confidence restriction on recollected items is supported by empirical data showing that recollection-based discrimination is reflected almost exclusively in high confidence ratings (Yonelinas, 1994, 2001). To the extent that this assumption is violated, the ROC would be more curvilinear. Consequently, the contribution of familiarity would be overestimated, and that of recollection would be underestimated.

2. The model also bears a superficial resemblance to two-criterion signal detection models of remembering and knowing in recognition (Donaldson, 1996; Hirshman & Henzler, 1998; Hirshman & Master, 1997; Inoue & Bellezza, 1998). However, the remember/know models assume that participants use an upper remember/know criterion to divide a single distribution of studied items into remember and know regions. In the present model, the upper criterion is used to discriminate between two separate distributions of studied items.

3. This constraint was included because most signal detection models of recognition assume that the variance of studied-item distribution does not normally decrease relative to that of the new-item distribution (i.e., the z-ROC does not have a slope greater than 1). For completeness, we fit a UVSD model with no range restrictions on S_s and S_w. The fitted weak-item distribution had a d'_w of 0.69 and an S_w of 0.75. This suggests that making sex judgments about faces during study increased the mean of the distribution but decreased the standard deviation. The overall R^2 was .979, which did not represent an appreciable difference from the R^2 of the UVSD model presented here with S range restrictions ($R^2 = .976$).

4. The ROC fitting procedure can also be conducted on raw response frequencies rather than cumulated frequencies. Maximum likelihood

estimation (Ashby, 1992; Ogilvie & Creelman, 1968) is sometimes preferred over least squares estimation because variability in both hits and false alarms can be accounted for. However, this problem was overcome in the present study by minimizing squared error terms for all item types. Both procedures, as well as the chi-square minimization technique employed by Slotnick et al. (2000), were performed on the present data, and all three yielded similar results. The maximum likelihood fits led to slightly lower R estimates and slightly higher d' estimates than did the least squares or the chi-square minimization of cumulated data.

However, the conclusions based on model comparisons using all the procedures were identical. Consequently, only least squares fits are reported for models (R^2), and F ratios are presented for comparisons among models.