JOURNAL OF MEMORY AND LANGUAGE **34**, 821–835 (1995)

Incorporating Response Bias in a Dual-Process Theory of Memory

ANDREW P. YONELINAS

University of California, Davis

GLENN REGEHR

University of Toronto, Toronto, Ontario, Canada

AND

LARRY L. JACOBY

University of Texas at Austin

We examined several different methods for incorporating response bias into a dual-process theory of recognition memory. Two high threshold correction methods, which have previously been applied to the process dissociation procedure, and a new, dual-process signal-detection method, were assessed. An examination of receiver operating characteristics (ROCs) showed that the threshold methods were inappropriate, but that the signal-detection method provided a reasonable account of the observed ROCs. Applying the corrections to a second data set showed that the different correction methods led to dramatically different conclusions, demonstrating that selecting the correct correction method is critical. Moreover, in agreement with the ROC analysis, the signal-detection method was the only one to provide a reasonable account of the data.

Dual-process theories of memory postulate that there are two qualitatively different processes or systems that support memory judgments. Evidence in support of these theories comes from observed dissociations between performance on indirect and direct tests of memory (for reviews see Richardson-Klavehn, & Bjork, 1988; Roediger & McDermott, 1993) and from dissociative effects on the processes that underlie recognition memory (e.g., Atkinson & Juola, 1974; Jacoby, 1991; Gardiner & Java, 1993; Mandler, 1980; Piercy & Huppert, 1972; Verfaellie & Treadwell, 1993; Yonelinas, 1994; Yonelinas & Jacoby, 1994). A question that arises when considering dual-process theories is how to incorporate response bias or guessing into such models. Differences in the response bias between different tasks make the interpretation of task dissociations difficult (Reingold & Merikle, 1991). Further, as we will show, the same differences also present problems for procedures designed to estimate the contribution of the underlying processes. Although numerous correction methods have been developed for single-factor models of recognition (see Murdock, 1974), there has been little discussion of these techniques within the framework of a dual-process theory.

In this paper, we begin by discussing the process dissociation procedure (Jacoby, 1991) which aims at estimating the contribution of recollection and familiarity to overall memory performance. We then examine several different methods that have been applied to that procedure to account for response bias. We first assess the different methods by examining the models that underlie each method and by contrasting the receiver operating characteristics (ROCs) predicted by the models against those observed in experimental data. We then assess the procedures further by applying them to a

Correspondence and reprint requests should be addressed to Andrew Yonelinas, Department of Psychology, University of California, Davis, California 95616. E-mail: apyonelinas@vcdavis.edu

0749-596X/95 $12.00

data set for which sizable response bias problems exist. The application of the correction procedures to that data set shows that distinguishing among these methods is critical because the different methods lead to dramatically different conclusions. The results of those analyses along with the ROC data show that several correction procedures that have been used in the past are insufficient, but that a dual-process signal-detection method provides a reasonable account of the data.

The Process Dissociation Procedure

In tests of recognition memory, subjects must decide whether a test item was presented in a previously studied list. Subjects could base recognition judgments on assessments of familiarity because, on average, an item that was presented in the study would be more familiar than one that was not. However, subjects may also be able to recollect some aspect of the study event (e.g., "I remember seeing the word . . . it was the first word in the list") and use this as a basis for recognition judgments.

Recollection and familiarity are said to differ in a number of ways. For example, familiarity is thought to be a fast basis for responding (Atkinson & Juola, 1974; Jacoby, 1991; Mandler, 1980) that relies on perceptual characteristics (Jacoby & Dallas, 1981) and reflects the automatic or unconscious use of memory (Jacoby, 1991) that is largely spared by amnesia (Piercy & Huppert, 1972; Verfaellie & Treadwell, 1993). In contrast, the use of recollection is described as a slow, search-like process that relies on conceptual processing or associative information and requires attention. Furthermore, recollection is said to be absent or reduced in amnesic patients.

Jacoby (1991) developed the process dissociation procedure to derive quantitative estimates for the contribution of recollection and familiarity to recognition memory performance. Estimates for recollection and familiarity were gained by contrasting performance in an inclusion condition where both processes acted in concert, to performance in an exclusion condition where the two processes acted in opposition. Although the procedure has been implemented in several different ways in recognition memory (see Yonelinas & Jacoby, 1994), let us consider the procedure as it was first used by Jacoby (1991). In phase 1 of that study, subjects read a list of words under incidental encoding conditions. In phase 2, subjects heard a different list of words and were instructed to remember them for a later recognition test. At test, subjects were presented with a list containing a mixture of words that were earlier seen, earlier heard, or new, and were given either *inclusion* or *exclusion* instructions. In the inclusion condition, they were instructed to call a word old if it was in either the seen or heard list. In the exclusion condition, they were instructed to call a word old only if it was in the heard list. Further, they were told that if they could recollect that the word was seen they could be sure the word was not heard, and they should call it new. That is, subjects were instructed to include seen words in the inclusion condition and exclude those words in the exclusion condition.

Performance in the inclusion and exclusion conditions for the *seen* words was used to derive estimates for recollection and familiarity. If the two processes are independent, the probability of responding "yes" to a seen word in the inclusion condition can be written as

$$P(\text{"yes"}/\text{old})\text{inc} = R + F - RF,$$

the probability that the item is recollected (R) plus the probability that it is familiar (F), minus the probability that the item is recollected and familiar (RF). That is, a seen item can be accepted as old if it is recollected as having been seen, or if it is sufficiently familiar to be judged old.

The probability of responding "yes" to a seen item in the exclusion condition can be written as

$$P(\text{"yes"}/\text{old})\text{exc} = F - RF,$$

the probability the item is familiar, minus the probability it is familiar and recollected. That is, subjects will only accept a seen word if it is familiar but they cannot recollect that it was seen. Recollection was calculated by subtracting the exclusion score from the inclusion score

[R + F − RF − (F − RF) = R]. Having solved for R, either of the two equations could be used to solve for familiarity [e.g., exclusion/(1 − R) = F].

Jacoby (1991) found that dividing attention selectively decreased recollection, and that words solved as anagrams led to greater recollection and familiarity than those simply read at study. Verfaellie and Treadwell (1993) used the same procedure and found that amnesia dramatically reduced recollection but left familiarity intact. A similar, although less dramatic effect, was found with aging (Jennings and Jacoby, 1994). Other variables such as increasing list length and decreasing retrieval time by speeding responses are found to reduce recollection but to leave familiarity intact (Yonelinas & Jacoby, 1994). In another study, which we return to later, relaxing response criterion was found to lead to an increase in the proportion of items accepted on the basis of familiarity but to leave recollection relatively unchanged (Yonelinas, 1994).

The validity of the estimates rely on two critical assumptions. The first assumption is that the two processes are independent. This assumption has been examined at length elsewhere (see Jacoby, Toth, & Yonelinas, 1993; Jacoby, Toth, Yonelinas, & Debner, 1994; Jacoby, Yonelinas, & Jennings, in press; Joordens & Merikle, 1993). Here, we focus on a second assumption which is that subject's criterion for responding does not differ in the inclusion and exclusion conditions. It is possible that under some conditions subjects are more lax with their use of familiarity in the inclusion condition, where both recollection and familiarity lead to correct responses, than they would be in the exclusion condition, where familiarity can lead to an incorrect response. If subjects do use different criteria for responding in the two conditions then this should be reflected in different false alarm rates to new items in the two conditions. Although in the past we have been careful to avoid such differences, differences do occasionally arise.

How can we solve for R and F when the base rates differ between inclusion and exclusion conditions? The solution involves incorporating the false alarm rates into the inclusion and exclusion equations. The aim of the current paper is to examine several different methods for incorporating false alarms in the process dissociation procedure and to assess their validity.

HIGH THRESHOLD CORRECTION METHODS

Several different high threshold correction methods have been used to measure the effects of memory and guessing within a single-factor theory of recognition memory. The general idea is that subjects can correctly identify an old item (Hit) either on the basis of "true" memory (i.e., a studied item exceeds a memory threshold) or on the basis of a guess, and that false alarms to new items (FA) arise because of guessing. The false alarm rate is used to estimate the probability of a guess, and guessing is algebraically removed from the hit rate to obtain a pure measure of memory. Although the procedures are typically used to produce a unitary measure of memory, we consider two such methods which have been applied to the process dissociation procedure.

HIT − FA. Probably the simplest way of correcting for false alarms is to subtract the probability of a false alarm from the probability of a hit. This is the correction method proposed for the process dissociation procedure by Roediger & McDermott (1994). They examined data from a study by Verfaellie and Treadwell (1993), which we mentioned earlier, in which the process dissociation procedure was used to examine the effect of amnesia on recollection and familiarity. Roediger and McDermott pointed out that there were differences in base rate between normal and amnesic patients. They argued that any conclusions drawn from that study can be questioned because the base rate differences may distort the estimates of recollection and familiarity. They went on to suggest a correction procedure in which base rates were subtracted from inclusion and exclusion scores before the estimates of recollection and familiarity were calculated. Although Verfaellie (1994) showed that the conclusions of the Verfaellie and Treadwell study were not changed by introducing such a correction procedure, we later discuss a study for which the correction procedure does influence conclusions.

The basic idea underlying the HIT-FA correction procedure is that the probability of a hit reflects true memory (recollection and familiarity) plus the probability that the subject correctly guesses that the item was studied. We can either subtract false alarm rates from the inclusion and exclusion scores before we use the process dissociation equations or we can add guessing terms to the inclusion and exclusion equations such that

$$P(\text{"yes"/old})\text{inc} = R + F - RF + Gi$$

$$P(\text{"yes"/old})\text{exc} = F - RF + Ge.$$

The proportions of new items accepted under inclusion and exclusion conditions are used as measures of guessing (Gi and Ge, respectively). Given the probability of accepting old and new items in the inclusion and exclusion conditions we can then solve the equations to derive estimates of R and F.

(HIT-FA)/(1-FA). Another common high threshold correction method is the (HIT-FA)/(1-FA) method. By this method the false alarm rate is first subtracted from the hit rate, then the sum is divided by one minus the false alarm rate. We later return to a study by Komatsu, Graf, and Uttl (1994) in which this correction procedure was applied to data from the process dissociation procedure. This correction method can be represented by writing the inclusion and exclusion equations in the following way:

$$P(\text{"yes"/old})\text{inc} = R + F - RF + Gi - Gi(R + F - RF)$$

$$P(\text{"yes"/old})\text{exc} = F - RF + Ge - Ge(F - RF).$$

As with the previous correction method, R and F can be calculated on the basis of the probability of accepting old and new items in the inclusion and exclusion conditions.

The notion behind this correction, at least as it is used in a single factor theories, is that guessing is independent of memory. This can be seen in the inclusion equation where overall performance is equal to memory (R + F − RF) plus guessing (Gi) minus the intersect (Gi(R + F − RF)). However, when the same correction is applied to the exclusion equation, guessing is no longer treated as being independent of memory (R + F − RF) but is treated as independent of (F − RF). In the Discussion, we return to a procedure that treats guessing as independent of memory in both equations.

Familiarity as a Signal-Detection Process

An alternative approach for incorporating false alarms into a dual-process model was proposed by Jacoby, Toth and Yonelinas (1993) who argued that false alarms arise because new items are incorrectly accepted on the basis of familiarity. That is, some new words may be familiar because of their preexperimental history, and subjects may incorrectly attribute this to the word's occurrence in the study list. They argued that false alarms should not be corrected from overall performance, which reflects recollection and familiarity, but rather from the probability of accepting an item on the basis of familiarity alone. Further, they suggested that the familiarity process might be described in terms of signal-detection theory.

The notion was elaborated by Yonelinas (1994), who argued that overall recognition performance reflected the independent contribution of a discrete recollection process, and familiarity process that reflects a Gaussian equal-variance signal-detection model. The idea is that all items have some level of preexperimental familiarity which can be described by a Gaussian distribution (see Fig. 1). Studying a list of items temporarily increases the familiarity of those items, which has the effect of shifting the distribution to the right.

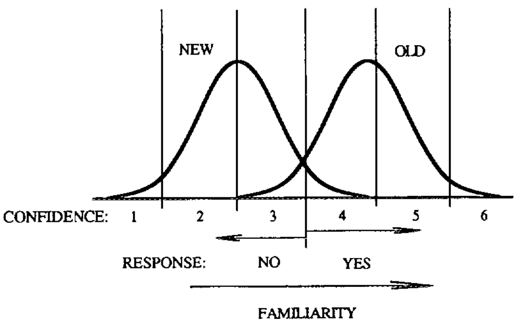

FIG. 1. Familiarity distributions representing old and new items in a Gaussian equal-variance signal-detection model.

The subject selects some level of familiarity so that only the items exceeding that level are judged as old. Independent of the familiarity process, subjects could correctly accept old items if they recollected some information about the study episode in which the item was encountered.

Important for this model and the associated correction method is that the response criterion is applied only to familiarity. Recollection is assumed to be independent of the false alarm rate, and false recollection of episodes which did not occur is assumed to be infrequent. So, as the criterion changes, the number of items accepted on the basis of familiarity will change but the probability of recollection will remain fixed. Later we discuss evidence in support of this assumption.

If familiarity is a signal-detection process, then the familiarity term in the original equations can be replaced by a function representing the probability that an old item exceeds the response criterion.

$$\Phi(d'/2 - C),$$

representing the proportion of the old item distribution exceeding the criterion (C) (see Macmillan & Creelman, 1991). When this term is substituted into the inclusion and exclusion equations we have

$$P(\text{"yes"}/\text{old})\text{inc} = R + \Phi(d'/2 - Ci) - R * \Phi(d'/2 - Ci)$$

$$P(\text{"yes"}/\text{old})\text{exc} = \Phi(d'/2 - Ce) - R * \Phi(d'/2 - Ce).$$

Of course, for a given value of C, there will be some proportion of new items incorrectly accepted as old. The false alarm rate will be equal to the proportion of the new item distribution exceeding the criterion, and this can be written as:

$$P(\text{"yes"}/\text{new})\text{inc} = \Phi(-d'/2 - Ci)$$

$$P(\text{"yes"}/\text{new})\text{exc} = \Phi(-d'/2 - Ce).$$

Thus we have four equations (the probability of accepting old and new items under inclusion and exclusion instructions) and four variables (R, d', Ci, and Ce). We can solve the equations to derive estimates for the four variables. However, because of the nature of the normal distributions that underlie the signal-detection model, a simple algebraic solution is not available. To solve for the unknowns we use a gradient descent search algorithm. The algorithm selects a set of parameter values (R, d', Ci, and Ce) and calculates a set of predicted values for the inclusion and exclusion scores and false alarms based on the above equations. The parameter space is systematically searched for the set of parameters that produces the observed scores by minimizing the sum of the squared differences between the observed and predicted values.[1] Alternatively, one could assume a logistic, rather than normal, distribution, in which case a clear algebraic solution does exist (see Appendix a). However, in keeping with the assumptions of signal-detection theory, we have assumed normal distributions and therefore must rely on the algorithm to locate the solution.

We should note that because familiarity is assumed to reflect a signal detection process, it will be measured in terms of d' rather than in terms of simple probabilities. This is because the probability that an item will be accepted on the basis of familiarity is not fixed and will vary with response criterion.

Given the three different correction methods, how do we assess which is most appropriate? In the current study, we first evaluate the correction methods by examining ROCs. Finally, the procedures are tested by examining the estimates they produce in a data set in which sizable base rate differences exist.

Assessing the Correction Methods: ROCs

ROCs provide a powerful tool for distinguishing between the correction methods because the models that underlie the three methods predict very different ROCs. By comparing the predicted and observed ROCs in the inclusion and exclusion conditions, we can assess each correction method.

An ROC is the function that relates the proportion of hits to the proportion of false alarms. Figure 2 presents ROCs for inclusion and exclusion conditions in a recognition memory test. The ROCs are from "long" lists (30 items

[1] The algorithm is available on request.

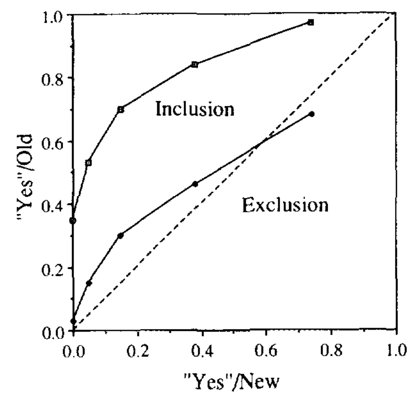

FIG. 2. Observed ROCs for inclusion and exclusion conditions for long lists (Yonelinas, Experiment 1, 1994).

in each) of Experiment 1 (Yonelinas, 1994). In that experiment, subjects studied two lists of words followed by a list discrimination yes–no recognition test. In one condition, subjects were instructed to respond "yes" if the item was in list 1, and in the other they were asked to response "yes" if the item was in list 2. Subjects were instructed to respond "no" if the item was new or if they could recollect that it was in the nontarget list. Thus, they should include items that they recollected from the target list and exclude items that they recollected from the inappropriate list. The inclusion score was measured as the probability of accepting a target word (i.e., list 1 words accepted under list 1? instructions and list 2 words accepted under list 2? instructions). The exclusion score was measured as the probability of accepting a nontarget item (i.e., list 1 words accepted under list 2? instructions and list 2 words accepted under list 1? instructions).[2]

Subjects made their responses on a 6-point confidence scale from *sure yes* (6) to *sure no* (1). ROCs were plotted as a function of response confidence such that the first point included only the most confidently remembered items (i.e., items eliciting a response of 6). The second point included all of the most confident responses plus the second most confident responses (i.e., items eliciting a response of 6 or 5). In this way the 6-point response scale produced a function with 5 points (see Fig. 2).

The inclusion function begins at .35 on the y axis and gradually increases in a curvilinear fashion toward the upper right corner of the graph. The curvilinear aspect of the ROCs does not seem to be a product of averaging, as the curve was seen across subjects, study positions, and test positions. The exclusion function begins at the x–y intercept and also increases in a curvilinear fashion, but remains far below the inclusion curve. Moreover, as the false alarm rate increases, the exclusion function crosses the diagonal. Thus, as response criterion is relaxed, the probability of accepting an old item in the exclusion condition drops below that of accepting a new item. This pattern was observed in several experiments in the Yonelinas (1994) study. Moreover, cases where exclusion scores drop below the false alarm rate are quite common (Debner & Jacoby, 1994; Jacoby, 1994; Jennings & Jacoby, 1993; Komatsu, Graf & Uttl, 1994; Toth, Reingold & Jacoby, 1994; Yonelinas & Jacoby, 1994).

Predicted ROCs for each correction method were generated by setting some level of recollection and familiarity, and varying the guessing or response bias parameter. This is analogous to examining performance as a function of response confidence. For the threshold models, G was varied from 0 to 1 (or until the hit rate was equal to 1.0) and the predicted inclusion and exclusion scores were plotted. For the dual-process signal-detection model, response criterion was varied from its minimum to its maximum value, and again inclusion and exclusion scores were plotted. Hypothetical ROCs for each correction method are presented in Fig. 3.

How did the predicted and observed ROCs compare? Examination of Figs. 2 and 3 shows that both of the threshold correction methods provided very poor fits to the observed ROCs. In contrast to the observed data they both predict linear (straight line) ROCs. Because the inclusion and exclusion equations are linear, increases in G produce proportional increases in the hits and false alarms, leading to straight line

[2] The inclusion condition in list discrimination procedure differs from the standard inclusion condition (i.e., accept items from either list): However, ROCs under list discrimination instructions were not found to differ significantly from those under standard inclusion instructions (Yonelinas, 1994).

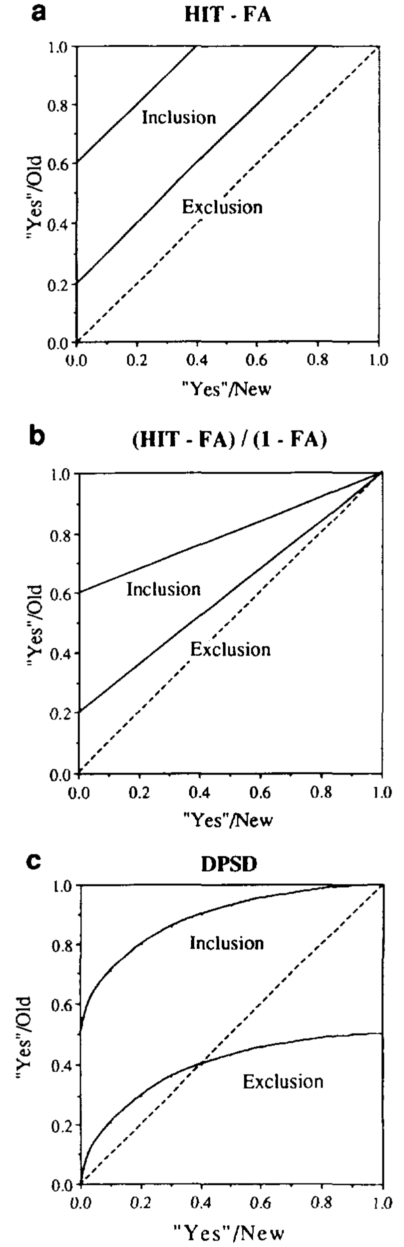

FIG. 3. Predicted ROCs for the (a) HIT − FA, (b) (HIT − FA)/(1 − FA), and (c) dual-process signal-detection models.

ROCs. This is particularly problematic for these correction methods because not only are curvilinear ROCs found under inclusion and exclusion conditions but they are almost always observed in standard studies of recognition memory (e.g., Donaldson & Murdock, 1968; Gehring, Toglia & Kimble, 1976; Glanzer & Adams, 1990; Ratcliff, Sheu, & Gronlund, 1992). In fact, the observed nonlinearity of ROCs was one of the primary reasons that threshold models of recognition were initially rejected as viable models for simple recognition judgments.

A second problem for the threshold correction methods is that neither can account for the exclusion function crossing the diagonal. If recollection and familiarity are greater than 0 then the probability of accepting an old item in the exclusion condition can never drop below the probability of a guess. This is because in the exclusion equations, guessing adds to the effects of memory. If R and F are greater than 0, then memory can only increase the probability of accepting items in the exclusion condition above the probability of a guess (the base rate). In fact, if the observed exclusion score drops below the base rate, the threshold correction methods would lead to negative estimates of familiarity. As we will see, this problem regularly arises when the threshold corrections are applied to experimental data.

In contrast to the threshold correction methods, the dual-process signal-detection method accounts well for the observed ROCs. In agreement with the data, the model predicts curvilinear functions for the inclusion and exclusion conditions. The reason for this is that the normal distributions that underlie the familiarity component lead to nonlinear increases in the probability of accepting old and new items. The correction procedure also accounts for the exclusion curve dropping below the diagonal. Because recollection of studied items leads to a "no" response in the exclusion condition, recollection places a ceiling on the number of old items that will receive a "yes" response. Thus the probability of accepting an old item in the exclusion condition should not increase beyond 1 − R. In contrast, because we assume that subjects do not recollect new items, the false alarm rate can increase to 1.0. Thus as the response criterion is relaxed and the false alarm rate increases toward 1.0, the exclusion score approaches 1 − R.

Although the signal-detection correction procedure does provide a better fit than either of the threshold corrections, we have found that the observed ROCs tend to deviate slightly from the predicted curves at the extreme levels of response confidence. The leftmost point on the inclusion function tends to be slightly lower than the model predicts. This discrepancy was discussed by Yonelinas (1994) and seems to arise either because some small proportion of recollected items did not lead to the most confident recognition judgments or because subjects were occasionally making list discrimination judgments based on the assessment of familiarity. In contrast, the right most point on the exclusion function tends to be slightly higher than the model predicts. This deviation may arise because of noise. If subjects were to occasionally respond randomly (e.g., accidentally hit the wrong key or simply guess) this would tend to force both the inclusion and the exclusion curves toward the diagonal (i.e., chance performance). We have found that, because the ROCs are cumulative, the effect of adding noise to the exclusion function is most pronounced as the false alarm rate increases. Thus, if subjects are occasionally responding randomly, the observed exclusion ROC will tend to rise above the predicted ROC as the response criterion becomes more lax. In any case, the deviations from the predicted curves are relatively minor, and are limited to the extreme high and low levels of response criterion. Our experience is that cases in which the correction method is required do not involve such extremes but rather fall in the middle of the functions where the model fits the data quite well.

In summary, despite some minor deviations at the extreme levels of response confidence, the dual-process signal-detection model predicted ROCs that were very close to those observed. It accounted for the curvilinearity of the inclusion and exclusion curves as well as the observation that the exclusion function dropped below the diagonal. The threshold corrections did not account for either of these findings.

Examining the Effects of Levels of Processing and Word Frequency

Although the dual-process signal-detection method provides a better account of the ROC data than either of the threshold models, one might ask if the difference between the correction methods is really that great and whether using the incorrect method would radically alter the conclusions drawn from a study. In the current section we examine a data set from Komatsu et al. (1994), in which sizable base rate differences were observed, and show that the different correction methods can lead to dramatically different conclusions. Let us consider that study in more detail.

Komatsu et al. (Experiment 1, 1994) used the process dissociation procedure to examine the effects of word frequency and levels of processing on recollection and familiarity in recognition memory. Subjects encoded a list of written words by answering questions about the pleasantness of the words (deep processing) or the number of syllables in each word (shallow processing). Half of the words in each condition were high-frequency and half were low-frequency words. Subjects then heard a list of words which they were to try to remember for a later memory test. Finally, subjects were given a recognition memory test under inclusion instructions (i.e., respond yes to words that were seen or heard), or exclusion instructions (i.e., respond yes only to words that were heard). The test list contained a mix of seen, heard, and new words. The probability of accepting previously seen high and low frequency words encoded under deep or shallow conditions and tested under inclusion or exclusion instructions are presented in Table 1. The probability of accepting new high and low frequency words are also presented.

An examination of Table 1 shows that the probability of accepting new items was not constant across conditions. Subjects were more likely to incorrectly accept new high frequency words than new low frequency words. Moreover, subjects accepted more high frequency words under inclusion instructions than under exclusion instructions.

Given these differences in base rate, Komatsu et al. (1994) realized they could not use the standard process dissociation equations, and echoed Roediger and McDermott (1994) by using correction procedures that correct for

TABLE 1
THE PROPORTION OF OLD AND NEW AND HIGH AND LOW FREQUENCY WORDS ACCEPTED UNDER INCLUSION AND EXCLUSION CONDITIONS

Encoding conditions	Word frequency	Test condition Inclusion	Exclusion
Pleasantness	Low	.89	.16
	High	.91	.37
Syllables	Low	.73	.25
	High	.74	.44
New	Low	.15	.16
	High	.38	.29

Note. Data from Experiment 1 in Komatsu et al. (1994).

differences in base rates before calculating estimates for recollection and familiarity. They used both the HIT-FA and the (HIT-FA)/(1-FA) correction methods. Table 2 presents those estimates along with estimates we calculated using the dual-process signal-detection method.

All three correction methods led to the same conclusions with respect to the effect of levels of processing and word frequency on recollection. Deeper processing led to an increase in recollection, and low frequency words were more likely to be recollected than were high frequency words. However, the patterns differed dramatically with respect to the effects on familiarity.

By the HIT-FA and the (HIT-FA)/(1-FA) methods, deeper processing led to lower estimates of familiarity than did shallow processing. Moreover, high frequency words led to greater increases in familiarity between study and test than did low frequency words. In contrast to the threshold corrections, the signal-detection method showed deeper levels of processing to increase familiarity and low frequency words to have a larger increase in familiarity than did high frequency words.

Clearly, the different correction procedures do lead to different conclusions, demonstrating that selecting the valid correction procedure is critical. Although the previous ROC analysis showed the threshold corrections to be inappropriate for the process dissociation procedure, can we reject any of these methods on the basis of the current data alone? In fact, an examination of the results based on the different correction methods strongly suggests that the two threshold correction can be rejected.

Consider the results produced by the high threshold models. Both the HIT-FA and the (HIT-FA)/(1-FA) correction methods led to the conclusion that familiarity decreased with deeper levels of processing. This "reversed" levels of processing effect conflicts with prior work using the process dissociation procedure, results from amnesic patients, and results from indirect tests. When levels of processing is examined under conditions where base rate differences were avoided (Toth, in preparation) the process dissociation procedure showed that deeper levels of processing led to an increase in both recollection and familiarity. Moreover, studies with amnesics suggest that deeper levels of processing does increase familiarity. Amnesics' recognition performance, which is presumably supported primarily by familiarity (see Piercy & Huppert, 1972; Verfaellie & Treadwell, 1993), is shown to benefit from deeper

TABLE 2
PARAMETER ESTIMATES FOR THE HIT-FA, THE (HIT-FA)/(1-FA) (FROM KOMATSU ET AL., 1994), AND THE DUAL-PROCESS SIGNAL-DETECTION (DPSD) CORRECTION METHOD FOR DEEP AND SHALLOWLY ENCODING, HIGH AND LOW FREQUENCY WORDS.

		HIT-FA		HIT-FA/(1-FA)		DPSD	
Encoding	Frequency	R	F	R	F	R	F(d′)
Pleasantness	Low	.73	−.01	.91	−.40	.73	1.25
	High	.45	.14	.79	.30	.50	1.21
Syllables	Low	.48	.18	.60	.23	.49	0.97
	High	.21	.19	.43	.35	.23	0.73

Note. Estimates are in terms of probabilities, except for familiarity in the DPSD method which is measured in terms of d′.

levels of processing (e.g., Mayes, Meudell, Neary, 1980). Finally, in indirect tests such as stem and fragment completion, deeper levels of processing is found either to have no effect on performance or to lead to a slight increase in performance (for a review see Roediger & McDermott, 1993). It seems most likely that the reversed levels of processing effect associated with the threshold corrections was an artifact produced by those correction procedures.

The threshold correction procedures also led to problematic conclusions regarding the effects of word frequency. They showed that the familiarity of high frequency words increased more than that of low frequency words. If low frequency words have a lower base rate familiarity then high frequency words, as the base rates in the current study show, then they should benefit most from the study phase. In fact, in a process dissociation experiment where base rates differences were not a problem, Jacoby (in preparation) found that the familiarity of low frequency words increased more than that of high frequency words, suggesting that the advantage associated with high frequency words was also an artifact of the threshold correction methods.

A final problem for the high threshold correction methods is that estimates for familiarity are sometimes negative. For example, Komatsu et al. present estimates for familiarity as low as −.40. Although our calculation—based on inclusion and exclusion data in Komatsu et al.'s tables—the estimates are closer to zero, it is clear that the threshold corrections can lead to negative probability estimates. Similarly, Roediger and McDermott (1994) report negative estimates for familiarity for the Verfaellie and Treadwell data when they used a HIT-FA procedure. Although small negative probabilities might reflect low levels of familiarity along with measurement error, the larger negative values suggest that the threshold correction methods are simply inappropriate.

In contrast to the threshold correction methods, the signal-detection correction method produced a much more reasonable account of the data. In agreement with prior research, the procedure showed that both recollection and familiarity increased with deeper levels of processing, and that low frequency words exhibited a greater increase in familiarity than did high frequency words.

The current analysis of Komatsu's data shows that choosing the correct correction method for the process dissociation procedure is critical because different correction methods lead to dramatically different conclusions. Second, based on the conclusions drawn using each method, it is clear that the high threshold correction methods lead to several problematic conclusions, and that the signal-detection method is the only one to produce a pattern of results that converges with prior research.

Alternative Correction Methods

In the current paper we examined three correction methods that have been applied to the process dissociation procedure. There are of course, numerous other possible ways in which base rates could be incorporated into the procedure. However, we examined several additional methods and none produced ROCs that fit the observed ROCs as well as that of the dual-process signal-detection model. Next, we briefly discuss a number of those alternatives.

One possibility that we considered is that guessing is a process which is independent of both recollection and familiarity. If this is true, then the inclusion and exclusion equations can be written:

$$P(\text{"yes"/old})inc = R + F - RF + Gi - Gi(R + F - RF)$$

$$P(\text{"yes"/old})exc = F - RF + Ge - Ge(R + F - RF).$$

This correction is very close to the (HIT-FA)/(1-FA) correction. In fact the inclusion equations are identical, and it is only the exclusions equation that differs—by the addition of the second R term. This difference allows the exclusion ROC for this model to drop below the diagonal. That is, as G approaches 1.0 the exclusion function approaches 1-R. Although this allows the method to account for the observation that exclusion function does drop below the diagonal, the model is linear and in contrast to the observed data, predicts straight line ROCs.

Another related threshold model that we considered assumes that experimental familiarity is additive with preexperimental familiarity. The equations for this correction method can be written:

$$P(\text{"yes"}/\text{old})\text{inc} = R + (F + Gi) - R(F + Gi)$$

$$P(\text{"yes"}/\text{old})\text{exc} = (F + Ge) - R(F + Ge).$$

This method predicts ROCs that are similar to the independent guessing method. The functions are linear and the exclusivity function drops below the diagonal. We have found that this correction method often leads to estimates that are very close to those of the dual-process signal-detection model. The relationship between these two methods is the same as that between standard signal-detection theory and a HIT-FA correction in single factor models. However, again the linearity of the model rules it out as a viable correction method.

An alternative method that we considered was to use a standard signal-detection correction on the overall recognition scores before using the process dissociation procedure. For example, one could convert the inclusion and exclusion scores into d′ values. However, this correction also runs into several problems. First, the standard signal-detection model predicts ROCs that begin at 0,0 and increase toward 1,1 and that are symmetrical along the negative diagonal. An examination of Fig. 2 shows that the ROCs are not symmetrical along the diagonal and that they intersect the y axis. To allow the signal-detection method to account for the asymmetrical ROCs, one could vary the old to new variance ratio. This could produce asymmetrical inclusion ROCs but would still have difficulty explaining why the curve intersects the y axis at a point close to the estimate for recollection. Even more problematic for this correction method is the observation that the exclusion function crosses the diagonal. We have found no way to allow the model to account for the exclusion curve dropping below the diagonal, short of introducing a discrete recollection process.

A related approach is to apply signal-detection theory separately to both familiarity and recollection. That is, treat the two processes as separate dimensions in a multidimensional signal-detection model (see Macmillan & Creelman, 1991). Although this approach may be useful in some applications of the process dissociation procedure, the low false alarm rate associated with recollection makes it unnecessary in the recognition memory experiments that we have examined. For example, in the ROC experiment discussed earlier, the false alarm rate for the highest confidence responses was zero for every subject, and the hit rate was very close to the estimate for recollection, suggesting that recollection can be treated as a discrete retrieval process. We are currently exploring other tasks, using the process dissociation procedure, for which a multidimensional signal-detection model may be appropriate.

When Do We Need To Correct For Guessing?

Although we examined several cases where there was a need to introduce a correction for base rates, these procedures are often not needed. Let us consider when the correction procedure is needed and the consequences of failing to apply the correction.

If the base rates do not change across conditions, then the standard inclusion/exclusion equations can be used—a correction method is not necessary. This is by far the simplest case and should be preferred. However, we should note that the estimates for familiarity will be dependent on the false alarm rate. If a measure of familiarity that is independent of base rate is required then familiarity should be converted to a d′ measure. This can be done with standard d′ tables by looking up the d′ value associated with the false alarm rate and the estimate of familiarity as the hit rate.

If base rates differ between inclusion and exclusion conditions then the standard equations will lead to distortions in the absolute estimates of recollection and familiarity. The most common difference we have observed is that subjects are more lenient with familiarity in the inclusion than the exclusion conditions and thus the false alarm rate is greater in the inclusion conditions. To obtain true estimates of these processes, or to compare the estimate of familiarity to new items, requires introducing the correction method.

However, if the goal of the study is to examine the qualitative effect of a variable on recollection and familiarity, then such a correction may not be necessary. For example, imagine a study in which we were interested in the effects of study time on recollection and familiarity. In this experiment, subjects study a list containing a mix of fast and slow items. Imagine that at test subjects accept more new items in the inclusion than in the exclusion condition. The difference in base rates between inclusion and exclusion conditions would inflate the estimates of R and F. However, because the base rates are the same for fast and slow items, the difference would be expected to introduce the same distortion to the estimates of both types of item. So, although the absolute estimates would be distorted, the qualitative effect of the variable (study time) on recollection and familiarity would be preserved. Thus a base rate difference between inclusion and exclusion conditions does not always require a correction method if the goal of the study is to examine the qualitative effects of some variable on recollection and familiarity.

If base rates differ across some experimental variable, the correction method may also be avoided. Differences of this sort would not influence estimates of recollection but would influence estimates of familiarity. To compare estimates of familiarity it would be necessary to convert the measures of familiarity (in probabilities) into d′ values. To do this one can use the false alarm rate along with the familiarity estimate (as the hit rate) and simply use d′ look-up tables to determine the d′ value. With familiarity measured in terms of d′ we can compare familiarity estimates independent of false alarm rate.

If there are differences between inclusion and exclusion conditions and between experimental conditions, as was the case in the Komutsu et al. (1994) study, then there seems to be no short cut, and the dual-process signal-detection correction method is required.

Applying the Model to Other Memory Tasks

The process dissociation procedure has been applied to numerous tasks other than recognition memory. For example, the procedure has been used to examine the contribution of recollection and automatic influences of memory in stem completion tasks (e.g., Jacoby, Toth & Yonelinas, 1993). Although the same dual-process signal-detection correction method may be applied in other domains, we advise considerable caution when doing so, and until an understanding of the processes in the particular task is understood, we would recommend trying to avoid base rate differences if at all possible.

We have begun to examine the application of the correction procedure to stem completion performance; however, the application of the correction method in that domain is considerably more complex and may not be appropriate. ROC analysis of stem completion performance is made difficult because, unlike recognition, subjects do not have to respond to all items. In fact, the task relies on subjects failing to respond correctly to many of the test items. Because of this, we cannot derive ROCs in the way we did for recognition performance. Moreover, base rate differences in stem completion may arise for different reasons than a shift in response criterion. For example, they may reflect a generate/recognize strategy rather than a shift in response criterion (see Jacoby, Toth & Yonelinas, 1993). Finally, it is not clear that we are dealing with exactly the same processes in the different memory tasks. Although it is likely that similar processes support recollection in recognition and stem completion, the processes that lead an item to seem familiar in a recognition task may not be the same as those that lead a word to come to mind in a stem completion task. Even if similar processes do support performance in different tasks it is likely that the different task demands will influence how those process affect performance.

One domain in which the current model may be useful is in studies of source monitoring. In a typical source monitoring experiment, subjects study items from two different sources. They are then given a recognition test for which they must first distinguish old items from new distractor items, and then are asked to judge the source of recognized items. A problem for current theories of source monitoring is that they cannot account for the observed recognition ROCs. Batchelder and Riefer (1990) proposed several multinomial models of source monitoring that have been used extensively in studies of source monitoring. However, Kinchla (1994) showed that ROCs generated by these multinominal models were not in agreement with a large body of data on recognition memory. The difficulty is that the multinomial models are high-threshold models and so must predict linear ROCs, rather than the curvilinear ROCs such as those found in recognition memory experiments.

We are currently examining the possibility of modeling performance on standard source monitoring tasks with the dual-process signal-detection model discussed in the current paper. By such a model, initial recognition judgments are based on recollection and familiarity, but source judgments reflect recollection alone. If familiarity reflects a signal-detection process that is independent of a discrete recollection process, then one would expect to see the type of curvilinear ROCs that are so problematic for current source monitoring models.

Summary

In current study we examined several different methods for incorporating response bias into a dual-process theory of recognition memory. We showed that when using the process dissociation procedure to estimate the contribution of recollection and familiarity, that differences in response bias can arise, and that the way in which these differences are accounted for leads to dramatic differences in the conclusions drawn. Several correction methods that have previously been applied to the process dissociation procedure were found to be inadequate. They led to conclusions that conflicted with prior research and were not able to account for important aspects of the observed ROC data. A dual-process signal-detection model, on the other hand, led to conclusions that were in agreement with prior research, and the model underlying the procedure was found to provide a reasonable account of the observed ROC data.

Appendix

The closed-form solution for R and d'F given H_I, FA_I, H_E and FA_E and assuming a logistic function for the cumulative function of the Familiarity (signal detection) component.

The clasic single-factor solution for d', assuming a logistic function, is

$$d' = \ln\left[\frac{H(1-FA)}{FA(1-H)}\right] \quad (A1)$$

Therefore, if R = 0, then

$$H_I = H_E = F = \frac{FA \cdot e^{d'}}{1 + FA \cdot (e^{d'} - 1)} \quad (A2)$$

However, if R ≠ 0, then

$$H_I = R + (1 - R)\left(\frac{FA_I \cdot e^{d'}}{1 + FA_I\,(e^{d'} - 1)}\right) \quad (A3)$$

and

$$H_E = (1 - R)\left(\frac{FA_I \cdot e^{d'}}{1 + FA_E\,(e^{d'} - 1)}\right) \quad (A4)$$

Therefore

$$e^{d'} = \frac{(H_I - R)(1 - FA_I)}{FA_I\,(1 - H_I)} = \frac{H_E\,(1 - FA_E)}{FA_E\,(1 - R - H_E)} \quad (A5)$$

$$(H_I - R)(1 - FA_I) \cdot FA_E\,(1 - R - H_E) = H_E\,(1 - FA_E) \cdot FA_I\,(1 - H) \quad (A6)$$

$$R^2 + R(H_E - H_I - 1) + \Big[H_I\,(1 - H_E)$$

$$- H_E(1 - H_I) \cdot \frac{FA_I(1 - FA_E)}{FA_E(1 - FA_I)} \Bigg] = 0 \quad (A7)$$

The solution for R is generated from Eq. (A7)

$$R = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad (A8)$$

where

$$a = 1 \quad (A8.1)$$

$$b = (H_E - H_I - 1) \quad (A8.2)$$

$$c = H_I(1 - H_E) - H_E(1 - H_I) \cdot \frac{FA_I(1 - FA_E)}{FA_E(1 - FA_I)}. \quad (A8.3)$$

And the solution for d′F is generated from Eq. (A5)

$$d'F = 1n\left[\frac{(H_I - R)(1 - FA_I)}{FA_I(1 - H_I)}\right]$$

$$= 1n\left[\frac{H_E(1 - FA_E)}{FA_E(1 - R - H_E)}\right] \quad (A9)$$

Note: Equation (A8) indicates that there are two solutions for R (i.e., there is not a unique solution as predicted by the program). However, the larger solution for R is always larger than H_I and $(R+H_E)$ is always greater than 1. Therefore when using the larger solution for R both versions of Eq. (A5) lead to a negative value for $e^{d'}$, so $d'F$ in Eq. (A9) is undefined. Thus, while there are two solutions for R, only one of these values leads to a real solution for $d'F$.

References

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce & P. Suppes (Eds.), *Contemporary developments in mathematical psychology (Vol. 1): Learning, memory & thinking* (pp. 242–293). San Francisco: Freeman.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review,* **97,** 548–564.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **20,** 304–317.

Donaldson, W., & Murdock, B. B. (1968). Criterion change in continuous recognition memory. *Journal of Experimental Psychology,* **76,** 325–330.

Gardiner, J. M., & Java, R. I. (in press). Recognizing and remembering. In A. Collins, M. A. Conway, S. E. Gathercole, & P. E. Morris (Eds.). *Theories of memory.* Hillsdale, NJ: Erlbaum.

Gehring, R. E., Toglia, M. P. & Kimble, G. A. (1976). Recognition memory for words and pictures at short and long retention intervals. *Memory and Cognition,* **4,** 256–260.

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.

Humphreys, M. S., & Bain, J. D. (1983). Recognition memory: A cue and information analysis. *Memory and Cognition,* **11,** 583–600.

Huppert, F., & Piercy, M. (1976). Recognition memory in amnesic patients: Effects of temporal context and familiarity of material. *Cortex,* **12,** 3–20.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging,* **8,** 283–293.

Jacoby, L. L. (1994). Measuring recollection: Strategic vs. automatic influences of associative context. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance XV.* Cambridge, MA: Bradford.

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

Jacoby, L. L. *The effects of word frequency on recollection and familiarity.* Manuscript in preparation.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General,* **3,** 306–340.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General,* **2,** 1–16.

Jacoby, L. L., Toth, J. P., Yonelinas, A. P., & Debner, J. A. (in press). The relationship between conscious and unconscious influences: Independence or redundancy? *Journal of Experimental Psychology: General.*

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. (in press). The relationship between conscious and unconscious (automatic) influences: A declaration of independence. In Cohen & Schoolers (Eds.), *Scientific approaches to the question of consciousness: The 25th annual Carnegie symposium on cognition,* Hillsdale, NJ: Earlbaum.

Joordens, S., & Merickle, P. M. (1993). Independence or redundancy? Two models of conscious and unconscious influences. *Journal of Experimental Psychology: General*, **122,** 462–467.

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

KOMATSU, S., GRAF, P., & UTTL, B. (1994). Process dissociation procedure: Core assumptions fail, sometimes. *European Journal of Cognitive Psychology*.

MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A users guide*. New York: Cambridge Univ. Press.

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

MAYES, A., MEUDELL, P., & NEARY, D. (1980). Do amnesics adopt inefficient encoding strategies with faces and random shapes? *Neuropsychologia*, **18,** 527–540.

MERIKLE, P. M., & REINGOLD, E. M. (1991). Comparing direct (explicit) and indirect (implicit) measures to study unconscious memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17,** 224–233.

MURDOCK, B. B. (1974). *Human memory: Theory and data*. New York: Earlbaum.

PIERCY, M., & HUPPERT, F. A. (1972). Efficient recognition of pictures in organic amnesia. *Nature*, **240,** 564.

RATCLIFF, R., SHEU, C. F., & GRONLUND, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *3*, 518–535.

RICHARDSON-KLAVEHN, A., & BJORK, P. M. (1988). Measures of memory. *Annual Review of Psychology*, **39,** 475–543.

ROEDIGER, H. L., & MCDERMOTT, K. B. (1993). Implicit memory in normal human subjects. In H. Spinnler & F. Boller (Eds.), *Handbook of neuropsychology* (Vol. 8). Amsterdam: Elsevier.

ROEDIGER, H. L., & MCDERMOTT, K. B. (1994). The problem of differing false-alarm rates for the process dissociation procedure: Comment on Verfaellie and Treadwell (1993). *Neuropsychology*, **8,** 284–288.

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

TOTH, J. P. *Deadlines, levels, and automatic processes in recognition memory*. Manuscript in preparation.

TOTH, J. P., REINGOLD, E. M., & JACOBY, L. L. (1994). Towards a redefinition of implicit memory: Process dissociations following elaborative processing and self generation. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **20,** 290–303.

VERFAELLIE, M. (1994). A re-examination of recognition memory in amnesia: Reply to Roediger and McDermott. *Neuropsychology*, **8,** 289–292.

VERFAELLIE, M., & TREADWELL, J. R. (1993). The status of recognition memory in amnesia. *Neuropsychology*, **1,** 5–13.

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., & JACOBY, L. L. (1994). Dissociations of processes in recognition memory: effects of interference and response speed. *Canadian Journal of Experimental Psychology*, **48,** 516–534.

(Received January 9, 1995)
(Revision received April 7, 1995)