

An individual difference analysis of false recognition

TIMOTHY A. SALTHOUSE AND KAREN L. SIEDLECKI

University of Virginia

Two studies with moderately large samples of participants were conducted to examine correlates of false recognition. In Experiment 1 false recognition of words was found to be a robust and reliable phenomenon at the level of individuals, and the tendency to classify critical lures as old was more closely related to the correct classification of old items as old than to the incorrect classification of unrelated new items as old. False recognition was not significantly related to any of the cognitive abilities that were assessed, including episodic memory, or to other factors such as personality and chronic mood. In Experiment 2 these findings were extended to include dot pattern and face stimuli. Although measures of veridical memory were significantly correlated across the different types of stimulus material, false recognition rates only had modest and generally not significant correlations, which suggests that the tendency to produce false recognitions may be a task-specific characteristic of individuals.

The false recognition phenomenon is the finding that new items that are related to actually presented items are falsely recognized as old in a recognition test much more often than are unrelated new items. A popular procedure used to investigate this phenomenon has come to be known as the Deese–Roediger–McDermott (DRM) paradigm because it was introduced by Deese (1959) and Roediger and McDermott (1995). The basic task involves the presentation of a set of items (typically words) that are related to one another in some fashion, with the participant instructed to study them in order to be able to remember them later. In the recognition version of the procedure the test phase contains an intermixed list of old items, unrelated new items, and new items that are related to the old items (which are called critical lures).

Because there are three types of test items (i.e., old, new, and critical lures) and two types of responses (i.e., “old” and “new”), three different proportions can be derived. Hit rates (HRs) correspond to the proportion of responses of “old” to old items relative to the number of old items that were presented, false alarms to new items (FANs) correspond to the proportion of responses of “old” to new items relative to the number of new items presented, and false alarms to critical lures (FACLs) correspond to the proportion of responses of “old” to critical lure items relative to

the number of critical lure items presented. Two additional measures can be computed from these proportions. The HR–FAN difference often is used as an index of veridical memory or corrected recognition, and the FACL–FAN difference has been used both as an index of corrected false recognition and as an index of reliance on gist (Koutstaal & Schacter, 1997; Koutstaal, Schacter, & Brennan, 2001; Koutstaal et al., 2003; Norman & Schacter, 1997; Schacter, Israel, & Racine, 1999). The rationale for the latter interpretation is that both the FACL and FAN proportions represent “old” responses made to new stimuli, but they differ in that the items used in determining the FACL proportion are related to the presented items, and therefore they might reflect reliance on gist information in addition to a general tendency to respond “old” to new items. (It should be noted that the FACL–FAN difference is only an approximate estimate of the degree of reliance on gist because a person could notice the relationship between the items on the basis of gist information but not make an “old” response to the item because he or she realized that it was new. The FACL–FAN index therefore may be an underestimate of the actual sensitivity to gist for people with good item-specific information.)

At present there are two major interpretations of the false recognition phenomenon. One attributes false memory responses to activation of related items at the time of study, combined with a failure to discriminate between internally and externally activated items at the time of test (Roediger, Watson, McDermott, & Gallo, 2001). Another view postulates that false recognitions are a consequence of greater reliance on memory of general features or gist compared to memory for specific details (Reyna & Brainerd, 1995). An extensive literature has developed over the past 10 years investigating these interpretations and the conditions that affect the magnitude of the false recognition phenomenon.

However, only a few studies have examined the false recognition phenomenon from an individual difference perspective. This is unfortunate because the patterns of correlations with another variable can be informative about what the variable represents, in a manner analogous to how one can learn something about a person by determining who he or she does, and does not, associate with. For example, if a measure of false recognition is moderately correlated with variable X but is unrelated to variable Y, then one may infer that the false recognition phenomenon probably has greater similarity to the processes involved in X than to those involved in Y.

There are several desirable requirements for meaningful correlational analyses. One is a moderately large sample of participants spanning a wide range of scores on the variables. Although correlations can be computed with samples of any size, correlations based on small samples have large confidence intervals, and consequently estimates of the magnitude of the relations are not very precise. The range of scores also should not be

restricted, as may occur when college students from a narrow ability range serve as research participants, because this can attenuate the magnitude of the correlations.

A second prerequisite for meaningful interpretation of correlations is evidence of the reliability of the relevant measures at the level of individual differences. There is often confusion between what might be called robustness and reliability. A phenomenon can be robust if there are many replications of a significant finding, but it may not be reliable at the level of the individual. In fact, there can be an inverse relationship between robustness of a within-participant effect and measurement reliability because statistical significance is high when there is little variation across people in the magnitude of the effect (because this quantity is in the denominator of the ratio used to determine significance), but the lack of variance between people in the magnitude of the effect often is associated with low reliability.

When reliability information is available, a standard statistical formula can be used to estimate the correlation that would have been obtained if the variables had been measured with perfect reliability. However, this is an indirect procedure for coping with unreliability, and the estimates can be misleading when the observed reliabilities are very low. An alternative approach to dealing with the issue of less-than-perfect reliability is to examine correlations at the level of latent constructs that represent the reliable variance shared by different measures. Because only systematic variance can be shared, these latent constructs are theoretically free of measurement error and therefore can be considered perfectly reliable.

We are aware of only two reports of the reliability of false memory measures, and the results have been inconsistent. In a sample of college students, Blair, Lenton, and Hastie (2002) reported internal consistency (coefficient alpha) estimates of .61 and .69 across two sessions and a 2-week retest correlation of .76. However, Lovden (2003) found that a false memory measure from the DRM procedure had an estimated (coefficient alpha) reliability of only .15 in a sample of adults ranging from 20 to 80 years of age.

A third desirable feature of correlational analyses is a multivariate rather than a bivariate perspective. The reason is that most cognitive variables are positively correlated with one another, and therefore misleading conclusions might be reached if a researcher were to focus only on the simple correlation between two variables. For example, if the target variable is related to two constructs X and Y, then some of the relationship of the target variable with X may be attributable to the influence of Y and vice versa. However, if both X and Y are included in the analysis, then the unique relationship of X (or Y) on the target variable can be determined to represent the influence that is statistically independent of the influence of the other variable.

Some researchers have used a quasicorrelational design in which participants are selected from the extremes of the distribution on one variable (or on a composite of several variables), and then everyone in each extreme group is treated as equivalent on the variable (i.e., they are categorized as either “high-X” or “low-X”). Although this type of design resembles an experiment in which participants are randomly assigned to groups or conditions, it has a number of limitations. For example, the magnitude of the correlation between the grouping variable and the target variable can be distorted because the variability in the middle of the distribution is omitted, and potentially valuable information is lost when people who vary continuously on a variable are assigned to a discrete category. Furthermore, merely because a researcher classifies participants as high on variable X does not mean that they are not also high on variable Y, and therefore when only the relationship with X is examined, there is ambiguity about whether it is actually the most relevant variable contributing to any relationship that might be observed.

Three types of correlations are relevant in the present context: internal, parallel version, and external. Internal correlations are correlations between different measures from the same task. Of particular interest are the correlations between the false recognition measure (FACL) and the hit rate (HR) and false alarm rate (FAN) measures. One of the reasons why the false recognition phenomenon has been considered so intriguing is that the mean levels of false recognitions often are more similar to the hit rate level than to the false alarm level. However, a key question from an individual difference perspective is whether people who make a high proportion of false recognitions also tend to make a high proportion of false alarms or a high proportion of hits. That is, is the phenomenon caused by a tendency for people to respond “old” to all types of new items, in which case a moderate to strong correlation would be expected between false recognitions (FACL) and false alarms (FAN)? Alternatively, are the people who are more likely to make high false recognitions (FACL) also more likely to make more hits (HR), as might be expected if people who are better at encoding information from actually presented stimuli are also more likely to activate information about related but not presented items?

Only a few studies have reported this type of correlational information, and perhaps because of the small samples, the results have been inconsistent. To illustrate, Tun, Wingfield, Rosen, and Blanchard (1998) reported weak correlations of FACL and HR (ranging from .06 to .45), whereas Koutstaal et al. (2001) reported correlations ranging from .44 to .83. Neither of these studies reported correlations between FACL and FAN, but Blair et al. (2002) reported weak correlations of FACL with both FAN (.35 and .22) and HR (.20 and .15).

Parallel version correlations are those between false recognition measures derived from tasks involving different types of stimulus materials. Information of this type is relevant to the question of whether the false recognition phenomenon reflects a characteristic of an individual rather than a property specific to particular tasks and materials. That is, if there is a dimension of individual differences associated with the tendency to make false recognitions, then one would expect moderate correlations between the false recognition measures derived from different types of materials. Comparisons across different types of materials are meaningful only if the false memory phenomenon can be demonstrated with other stimulus materials, and there is evidence that this is the case. To illustrate, there are reports of higher false alarms to related (i.e., *FACL*) than to unrelated (i.e., *FAN*) new words with phonologically related words (Sommers & Huff, 2003) and with pictures or drawings from the same category compared to pictures or drawings from other categories (Intons-Peterson, Rocchi, West, McLellan, & Hackney, 1999; Koutsaal & Schacter, 1997; Koutsaal, Schacter, Galluccio, & Stofer, 1999; Koutsaal et al., 2001, 2003; Lovden, 2003). However, only the Lovden study reported correlations between the false memory measures derived from different types of material. The correlation between the false recall of critical lure words in the DRM paradigm and false recognition of new but related pictures was .21, but this value is difficult to interpret because the reliabilities of the measures were low, and there was no adjustment for the influences of veridical memory on the false memory measures.

External correlations are a third type of relevant correlation in which the correlation of interest is between the false recognition variable and other types of variables, such as the person's age, level of various cognitive abilities, personality traits, or mood. Age is an interesting variable in this context because memory is the aspect of cognitive functioning most often assumed to decline with increasing age, and therefore one might expect a measure of erroneous memory to be particularly susceptible to effects of aging. Surprisingly, however, research results generally have not been consistent with this expectation: There are numerous reports of no age differences in false recognitions in at least some conditions (Benjamin, 2001; Budson, Daffner, Desikan, & Schacter, 2000; Gallo & Roediger, 2003; Intons-Peterson et al., 1999; Kensinger & Schacter, 1999; Koutsaal et al., 2001, 2003; McCabe & Smith, 2002; Norman & Schacter, 1997; Schacter et al., 1999; Thomas & Sommers, 2005; Tun et al., 1998).

One factor that could contribute to the failure to find consistent age differences in false memory is that differences in veridical memory have not always been taken into account. Some researchers have dealt with this issue by attempting to match adults of different ages on a measure of veridical memory (Balota et al., 1999; Koutsaal et al., 2001; Sommers

& Huff, 2003), but the resulting samples often are small and have limited statistical power. Another procedure is to statistically remove the influence of veridical memory from the false memory variable before examining its relationship to age or other variables. This method has been used infrequently in research on age differences in false memory (Sommers & Huff, 2003; Watson, Balota, & Sergent-Marshall, 2001), possibly because the samples have been small and the estimates would have been imprecise.

Another type of external relationship consists of correlations of the false recognition measure with different cognitive abilities, including episodic memory. If false recognition is another manifestation of memory processing, then one might expect veridical recognition (HR-FAN) and false recognition (FACL) measures to have moderate correlations with episodic memory ability but weak correlations with other cognitive abilities. However, if false recognitions are a reflection of weak attentional control or limited working memory involvement (Balota et al., 1999; McCabe & Smith, 2002; Watson, Bunting, Poole, & Conway, 2005) or of poor inhibitory control (Lovden, 2003; Sommers & Huff, 2003), then one might expect stronger correlations of false recognition with measures of these other constructs.

Several studies have reported correlations of false memory with other cognitive variables, but the available results are not easily interpretable. For example, it is not clear why people who perform better on the Boston Naming Test (Balota et al., 1999; but not Sommers & Huff, 2003) or on an arithmetic test (Butler, McDaniel, Dornburg, Price, & Roediger, 2004) are less likely to make false memories. Furthermore, in some cases correlations of the other cognitive variables were significant for measures of both veridical and false memory, and therefore at least some of the correlations with false memory may not be independent of those on veridical memory (Butler et al., 2004; Lovden, 2003; Sommers & Huff, 2003; Watson et al., 2005). Finally, many of the studies have relied on single measures of the other constructs (McCabe & Smith, 2002; Sommers & Huff, 2003; Watson et al., 2005; but see Butler et al., 2004, and Lovden, 2003), which are less reliable, and more likely to contain construct-irrelevant influences, than latent constructs based on the variance common to several measures.

Correlations of false recognition with various noncognitive variables are also of interest because the tendency to respond "old" to related new items might be a manifestation of a personality characteristic such as extroversion or openness or attributes of mood such as anxiety or depression. To the extent that this is the case, significant correlations of false recognitions with these traits might be expected. The relationship between the tendency to make false memory responses and a measure

of dissociative tendencies has been examined, but one study (Winograd, Peluso, & Glover, 1998) found a significant correlation, and one (Wright, Startup, & Mathews, 2005) did not.

Two recent studies using a mood induction manipulation found opposite results regarding the effects of mood on the tendency to make false memory responses. Wright et al. (2005) reported that negative mood was associated with more false recalls when participants were asked to recall as many words as possible, but Storbeck and Clore (2005) reported lower levels of false recall among participants in the negative mood condition. Several differences in the procedure across the studies may be responsible for the different results because the mood induction in the Wright et al. study was after the presentation of the items, but it was before the presentation of items in the Storbeck and Clore study, and participants who did not exhibit the expected mood change were eliminated from the analyses in the Storbeck and Clore study, whereas Wright et al. kept all participants in the analysis and used liking of the music that was used to induce the mood state as the measure of mood. Although the available evidence concerning the relationship of false memory to acute mood state is inconsistent, it is possible that false recognition might be related to chronic mood as assessed with questionnaires concerned with mood, anxiety, depression, or aspects of personality such as neuroticism.

Results from two separate studies are reported, each consisting of independent samples of 327 adults spanning a wide age range. The first study involved only word stimuli but asked participants to make remember-know judgments in addition to old-new decisions. The second study involved words, faces, and dot pattern stimuli, as represented in Figure 1. Both studies relied on latent constructs to represent veridical recognition (HR-FAN), false recognition (FACL), and corrected false recognition or gist (FACL-FAN) in the correlational analyses.

All participants in each study also completed a battery of 16 cognitive tests (described in Salthouse, 2004, 2005) that were used to create latent constructs representing four distinct cognitive abilities: fluid intelligence, verbal episodic memory, perceptual speed, and vocabulary. Standardized coefficients for the construct-variable and construct-construct relationships obtained from confirmatory factor analyses are reported in the Appendix. The tests used to represent fluid intelligence involved selecting the best completion of a matrix containing geometric patterns in all but one cell, determining which set of letters differed from the others, complete a series of items with different types of materials, determining the correspondence between two-dimensional and three-dimensional drawings of objects, selecting which pattern of folds would result if a piece of paper were folded in a specified manner and a hole punched

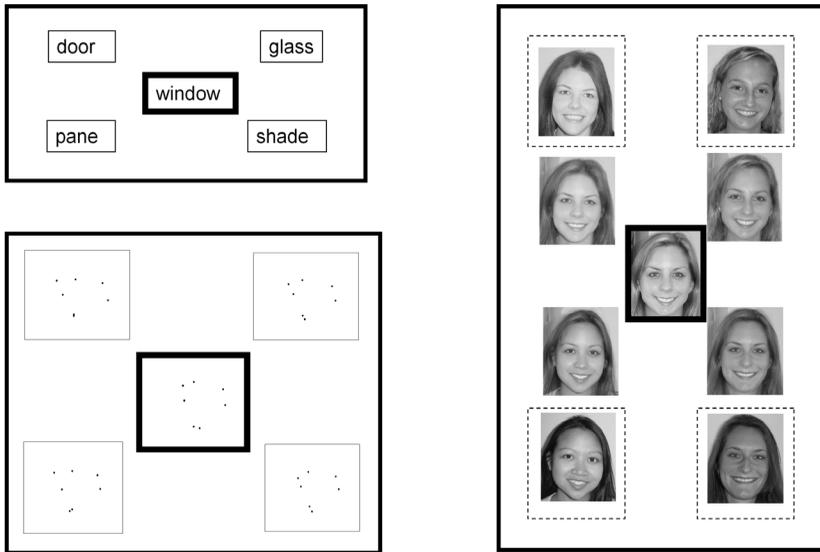


Figure 1. The three types of stimuli used in Experiment 2. In each case the item in the middle corresponds to the critical item that is related to the other items but is presented only at test

in a particular location, and determining which geometric shapes could be combined to produce a target pattern. The episodic memory tests consisted of multiple-trial free recall of words, recall of idea units from narratives, and recall of paired associates created from unrelated words. Perceptual speed was assessed with a Digit Symbol substitution test and tests of letter comparison and pattern comparison. Vocabulary was assessed with tests requiring that the participant define words, name pictured objects, or select the best synonym or the best antonym of the target word. Finally, all participants also completed questionnaires to assess aspects of personality, depression, anxiety, and positive and negative mood.

EXPERIMENT 1

The false memory task involved the participant viewing a set of words and then classifying each test item as old or new in an immediate recognition test. In addition, after each old–new decision the participants were asked to indicate whether the preceding decision was based on deliberate remembering or simply based on knowing the answer (Gardiner & Parkin, 1990). The purpose of this instruction was to determine whether the two types of responses differed in their patterns of correlations with other

variables, as would be expected if they represented qualitatively distinct types of processing.

METHOD

Participants

Complete data on the false memory task were available from 327 adults ranging from 18 to 93 years of age (mean = 50.7, standard deviation = 17.6). Three participants were deleted from the analyses because they scored less than 24 on the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) often used to screen for dementia. The participants were recruited from the community via newspaper advertisements, flyers, and referrals from other participants and were paid for their participation. Characteristics of the sample are summarized in Table 1, where it can be seen that higher age was associated with slightly poorer self-reported health but with somewhat higher age-adjusted scaled scores on vocabulary, Digit Symbol, Logical Memory, and Word Recall.

Procedure

The false memory task was administered in the second session of a project in which adults of different ages reported to the laboratory for three sessions during which they performed a variety of cognitive tasks and completed a number of questionnaires. Details of the specific tests are described elsewhere (Salthouse, 2004, 2005; Salthouse, Berish, & Siedlecki, 2004), and results of confirmatory factor analyses supporting the correspondence between variables and abilities

Table 1. Descriptive characteristics of sample, Experiment 1

	Age			Age <i>r</i>
	18–39	40–59	60–94	
<i>N</i>	89	131	105	NA
Age (years)	28.3 (6.0)	49.8 (5.4)	70.9 (7.7)	NA
Percentage female	66	66%	60%	-.04
Self-rated health	1.9 (0.8)	2.1 (0.9)	2.3 (0.9)	.21*
Education (years)	15.2 (2.2)	15.8 (2.8)	15.9 (2.7)	.13*
Scaled scores				
Vocabulary	12.8 (3.1)	12.2 (3.0)	13.6 (2.9)	.10
Digit Symbol	11.0 (2.8)	10.6 (2.9)	11.4 (2.7)	.09
Logical Memory	11.9 (2.5)	11.9 (2.3)	13.0 (2.8)	.16*
Word Recall	11.8 (3.2)	11.7 (3.4)	12.4 (3.3)	.08

Note. Values in parentheses are standard deviations. Health was a self-rating on a scale ranging from 1 (*excellent*) to 5 (*poor*). Scaled scores are age-adjusted scores relative to the nationally representative normative sample in the Wechsler Adult Intelligence Scale III (Wechsler, 1997a) and Wechsler Memory Scale III (Wechsler, 1997b), which have means of 10 and standard deviations of 3 in the normative sample.

* $p < .01$.

are summarized in the Appendix. Among the questionnaires completed by the participants were the Center for Epidemiologic Studies Depression Scale (Radloff, 1977), the Five Factor Personality Inventory (Goldberg, 1999), the Spielberger Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushere, 1970), and the Positive and Negative Affect Scale (Watson, Clark, & Tellegen, 1988).

The 10 lists of 15 words each were visually presented on a computer screen at a rate of 1 s per word, with 1 s between words and 2 s between lists. The stimulus words were obtained from the lists provided in Stadler, Roediger, and McDermott (1999). The top 20 lists in terms of false recognition of the critical lure were used, with items from the odd-numbered lists presented in the study phase and new items for the recognition test taken from the even-numbered lists. The recognition test, consisting of 30 old words, 10 critical lures, and 40 new words, was administered immediately after the last study item. It contained items from serial positions 4, 8, and 12 plus the critical lure from each of the 10 stimulus lists, as well as three items and the critical lure from 10 lists that were not presented in the study phase. Each successive set of eight items in the recognition test contained one new related item (i.e., the critical lure target item), three old items, and four unrelated new items.

Two responses were made to each test item. The first response was a press of the “Z” key to indicate whether the test word was old, in the sense that it had occurred in one of the previous lists, or a press of the “M” key to indicate that the word was new and had not been presented in any of the earlier lists. The second response indicated the basis for the first decision. If the old–new decision was based on remembering or conscious awareness of the presence or absence of the word in an earlier list, the “R” key was to be pressed. If the old–new decision was based on simply knowing, without a conscious recollection of its actual presence or absence, then the “K” key was to be pressed.

RESULTS

Means and standard deviations of the primary dependent measures across all trials in the false memory task are presented in the top row of Table 2. Inspection of the entries reveals that the mean for the FACL measure was greater than not only the FAN measure but also the HR measure. The FACL measure was greater than the FAN measure for 96% of the participants, and it was greater than the HR measure for 76% of the participants.

The second and third rows in Table 2 contain the response proportions for decisions attributed to remembering and for decisions attributed to knowing. All of the values were significantly greater for the “remember” decisions than for the “know” decisions (i.e., $t_s > 6.8$). As expected, veridical memory (HR–FAN) was higher with “remember” decisions than with “know” decisions. However, the FACL–FAN index sometimes used to reflect reliance on gist was also larger for “remember” than for “know”

Table 2. Summary statistics for memory parameters, Experiment 1

	HR	FACL	FAN	Correlations with FACL	
				HR	FAN
				Proportion "old"	
All items	.69 (.17)	.80 (.22)	.19 (.17)	.63*	.16*
"Remember"	.51 (.22)	.59 (.29)	.10 (.12)	.75*	.27*
"Know"	.18 (.18)	.21 (.23)	.09 (.12)	.81*	.25*
				Proportion "remembered"	
	.73 (.25)	.72 (.29)	.53 (.37)	.78*	.35*
				Decision time (s)	
	2.52 (1.07)	2.49 (1.66)	3.56 (1.91)	.72*	.52*

Note. Values in parentheses are standard deviations. FACL = false alarm to critical lures; FAN = false alarm to new items; HR = hit rate.

* $p < .01$.

decisions, which would not be expected if "know" decisions were based predominantly on gist information.

The fourth row in Table 2 contains the proportion of each type of response attributed to remembering. Once again, the mean values were fairly similar for the HR and FACL measures, and both were larger than the value for the FAN measure. The final row in Table 2 contains decision times in the recognition test. Because there was no explicit instruction to respond rapidly, these decision times probably represent a mixture of capability and stylistic influences. It is nevertheless noteworthy that the median times for the HR and FACL responses were very similar, and both were much shorter than the times for the FAN responses.

The results just described indicate that, at least in terms of the mean levels, FACL responses were more similar to HR responses than to FAN responses. The last two columns in Table 2 contain correlations that reveal that there was also a stronger correspondence in the ordering of participants between the FACL and HR measures than between the FACL and FAN measures. That is, the correlations indicate that people who make a high proportion of false recognitions are much more likely to be higher than average in their hit rates than in their false alarm rates. As noted earlier, this result is more consistent with the false recognition phenomenon originating because of processes at encoding than processes at the time of response.

Because the FACL measure was significantly correlated with the veridical memory (HR-FAN) measure (i.e., .38 for all trials, .59 for "remember" trials, and .66 for "know" trials), subsequent analyses were designed to take

the variation in true recognition into account when examining correlations with FACL. This was accomplished by creating residuals of the FACL measures after controlling the respective veridical memory measures and then repeating the correlational analyses on these residuals.

In order to maximize the reliability of the relevant memory measures, latent constructs for veridical recognition (HR-FAN), false recognition (FACL), and corrected false recognition/gist (FACL-FAN) were formed by using the scores on the two parts (i.e., the first 40 trials and the second 40 trials) of the test as separate indicators. The use of only two indicators of a construct is not ideal and in some cases can lead to problems of identification, but this can be avoided by constraining the residual variances of the indicators to be equal. Some of the other variables in the correlational analyses (e.g., age and the scale scores from the questionnaires) were represented by a single indicator. However, the cognitive abilities were represented as latent constructs in a structural equation model in which all abilities were considered simultaneously.

Correlations of the parameters from the false memory task with demographic variables, cognitive abilities, and personality variables are presented in Table 3. The veridical memory measure for all trials (HR-FAN) was higher for participants with better verbal episodic memory, and there was an independent correlation of vocabulary, indicating that people who know more word meanings were also more accurate at discriminating old from unrelated new items. There was a similar pattern of correlations for decisions attributed to remembering, although the vocabulary correlation was weaker and no longer significantly different from zero. The veridical memory measure for decisions attributed to knowing was not significantly related to any of the demographic, cognitive ability, personality, or mood variables.

The moderate correlations between the veridical memory measures and episodic memory ability are not surprising because both constructs represent memory for verbal materials. The lack of a correlation with the measure derived from the “know” trials is consistent with the interpretation that knowing decisions reflect a qualitatively different type of processing than the deliberate and effortful processing involved in many episodic memory tasks.

The correlations of greatest interest are those with the false recognition (FACL) measure, both before and after differences in veridical recognition are adjusted for. None of the correlations were significant in the data from all trials, and the significant effects with age on the two types of decisions were opposite in direction: Higher age was associated with more false recognitions for decisions attributed to remembering but with fewer false recognitions for decisions attributed to knowing.

Correlations involving the false recognition measure were expected to

Table 3. Correlations of the memory parameters, Experiment 1

	All						"Remember"						"Know"					
	HR-FAN		FACL-FAN		FACL		HR-FAN		FACL-FAN		FACL		HR-FAN		FACL-FAN		FACL	
Age	.05	.09	.12	(.15)	.08	.19*	.24* (.24*)	-.13	-.14*	-.20* (-.16*)								
Sex (male = 0)	.13	.14	.14	(.13)	.14	.15*	.11 (.03)	-.03	-.04	-.02 (.00)								
Education	.15	.07	-.04	(-.08)	.05	-.01	-.08 (-.14*)	.10	.09	.09 (.03)								
Fluid intelligence	-.10	-.01	-.19	(-.12)	-.23	-.23	-.26 (-.15)	.16	.26	.18 (.09)								
Memory	.43*	.26	.10	(-.01)	.40*	.28*	.16 (-.10)	-.05	-.09	-.09 (-.07)								
Perceptual speed	.01	-.13	-.03	(.01)	-.03	-.11	-.12 (-.13)	.07	.00	.11 (.09)								
Vocabulary	.21*	.25*	.18	(.05)	.15	.17	.11 (.03)	.10	.03	-.01 (-.10)								
Neuroticism	-.13	-.16	-.12	(-.09)	-.12	-.09	-.09 (-.02)	-.01	-.04	.03 (.04)								
Extroversion	-.02	-.06	-.09	(-.09)	-.02	-.02	-.05 (-.05)	-.01	-.03	.01 (.02)								
Openness	.13	.15	.08	(.14)	.04	.06	.01 (-.01)	.10	.08	.09 (.03)								
Agreeableness	-.01	.01	.02	(-.03)	-.02	.02	.00 (.02)	-.02	-.02	.04 (.07)								
Conscientiousness	.03	.07	.02	(.05)	-.04	-.00	-.06 (-.05)	.09	.06	.09 (.04)								
Trait anxiety	-.12	-.19*	-.12	(-.10)	-.08	-.11	-.08 (-.04)	-.05	-.06	.00 (.05)								
Depression	-.12	-.17*	-.09	(-.06)	-.07	-.09	-.07 (-.04)	-.06	-.05	.04 (.10)								
PANAS, positive	.01	.04	.06	(.08)	-.05	.02	.06 (.11)	.05	.02	-.01 (-.06)								
PANAS, negative	-.06	-.15	-.09	(-.08)	-.07	-.09	-.04 (.00)	.01	-.04	-.01 (-.02)								

Note. Values in parentheses are standardized regression coefficients on residuals created by partialling of the influence of HR-FAN from the FACL measures. Because a sizable number of participants had few "remember" or "know" responses, a large proportion of data was missing when the performance measures were computed for "remember" and "know" responses on each half of the test. The "remember" and "know" correlations therefore are based on the average across all trials rather than on a latent construct defined by estimates from the first and second half of the test, as was the case for the values in the "All" columns. FACL = false alarm to critical lures; FAN = false alarm to new items; HR = hit rate; PANAS = Positive and Negative Affect Scale. * $p < .01$.

be informative about what the false recognition phenomenon represents. Surprisingly, the false recognition measure was not significantly related to episodic memory ability, indicating that people who performed well in other verbal episodic memory tasks were no more likely to make false recognitions than people who performed poorly. There were also no significant correlations of the FACL measure with other cognitive abilities, personality dimensions, or anxiety, depression, or level of positive or negative affect.

It is important to note that the lack of correlation with the false recognition measure is not attributable to weak reliability because the latent construct for false recognition was well defined with standardized regression coefficients of .72 and .76. Furthermore, a direct estimate of reliability (coefficient alpha) derived from the scores on the first and second halves of the test was .71.

The FACL–FAN measure sometimes used as a measure of corrected false recognition or as an index of reliance on gist was not related to age for all trials, but there was a positive age correlation for “remember” decisions and a negative age correlation for “know” decisions. Both of these correlations are largely attributable to correlations of age with the FACL measures because there were no significant correlations of age with the FAN measure. The negative correlations between the FACL–FAN index and anxiety and depression indicate that participants with higher levels of anxiety and depression had lower FACL–FAN values, suggesting that, to the extent that this index reflects gist processing, there may be less reliance on gist when people are anxious or depressed.

DISCUSSION

The results of this study indicate that the false recognition phenomenon is robust, in the sense that it is exhibited by most people, and is reliable, in the sense that it is consistent across different parts of the test. The findings are also unambiguous with respect to the question of whether the tendency to report that related items were old (i.e., FACL) is more closely related to the tendency to say “old” to unrelated new items (i.e., FAN) or to the tendency to say “old” to actual old items (i.e., HR). The comparisons of both means and correlations revealed that the FACL measure was more similar to the HR measure than to the FAN measure. This pattern of results is consistent with the interpretation that the false recognition phenomenon occurs because during the initial presentation the related new items are activated and function as old items at the time of the recognition decision, such that responses to old items and to critical lure items are nearly indistinguishable with respect to frequency, decision time, and attribution to remembering or knowing.

The analyses were less successful at identifying characteristics of people associated with the tendency to respond “old” to new related information. Unlike the HR-FAN (veridical memory) measure, the FACL measure was not significantly related to episodic memory ability or to any of the other cognitive abilities. Because the false recognition variable was measured reliably, these results imply that, from the perspective of individual differences, the tendency to falsely identify related new items as old reflects a different dimension of human functioning than what is captured by these cognitive abilities. Unfortunately, it is not clear what that dimension represents because there were also no correlations of this tendency with several personality dimensions or frequently used measures of mood.

EXPERIMENT 2

The discovery that false recognition could be measured reliably but was not correlated with other variables is surprising and raises the possibility that the false recognition phenomenon may be specific to a particular type of task rather than reflecting a characteristic of individuals that is manifested in a variety of similar situations. Experiment 2 therefore was designed to examine false recognition with three different types of materials to determine whether the same phenomenon is apparent with each type of material and, if so, to examine the degree to which the tendency to make false recognitions is correlated across types of material. If false recognition measures with different types of materials are correlated across individuals, and if they exhibit similar correlations with other variables, then the false recognition phenomenon could be inferred to represent a characteristic of the individual that is at least somewhat independent of the specific type of stimulus material. In contrast, if the measures are not related to one another and do not exhibit the same pattern of correlation with other variables, then the false recognition phenomenon may be highly task specific.

Figure 1 illustrates the three types of materials used in the study. In all cases the item in the center was closely related, either semantically or on the basis of physical resemblance, to the other items, but it was presented only in the test list and not in the study list. The task with word stimuli was identical to that used in the previous study except that the participants only made an “old” or “new” response and did not make a “remember” or “know” categorization. The dot pattern stimuli were loosely based on research by Posner and Keele (1968, 1970), who investigated artificial concepts by creating exemplars from prototype dot patterns and found that the prototypes frequently were falsely categorized as old in later memory tests. The face stimuli were created by morphing different original faces

to the same target face and then presenting only the morphed faces at study, but the target faces together with new and old morphed faces in the memory test. Similar to the Posner and Keele studies, a study by Solso and McCarthy (1981) using artificially created faces found that new prototype faces frequently were falsely recognized as old in a recognition test.

As in Experiment 1, participants reported to the laboratory for three sessions of approximately 2 hours each to complete the false memory tasks and a number of other tasks designed to measure a core set of cognitive abilities. Personality, depression, anxiety, and affect questionnaires were also completed by all participants.

METHOD

Participants

The participants were 332 people between the ages of 18 and 94 who were recruited in the same manner as in Experiment 1. Five participants were excluded from subsequent analyses because they scored less than 24 on the MMSE. Descriptive characteristics of the sample are presented in Table 4, where it can be seen that higher age was associated with slightly poorer self-reported health but a greater number of years of education. Age was also associated with higher scaled scores for the Digit Symbol and Logical Memory subtests of the Wechsler Adult Intelligence Scale III, which suggests that the older adults in the sample

Table 4. Descriptive characteristics of sample, Experiment 2

	Age			Age <i>r</i>
	18–39	40–59	60–94	
<i>N</i>	102	117	108	NA
Age (years)	26.3 (6.4)	50.8 (5.7)	71.3 (7.8)	NA
Percentage female	63	76	68	.05
Self-rated health	1.8 (0.8)	2.1 (0.8)	2.1 (0.8)	.17*
Education (years)	14.9 (2.3)	15.3 (2.8)	16.7 (3.6)	.26*
	Scaled scores			
Vocabulary	13.4 (3.4)	11.9 (3.2)	13.8 (2.6)	.06
Digit Symbol	10.5 (3.0)	10.9 (3.0)	11.5 (2.5)	.15*
Logical Memory	11.5 (2.8)	11.3 (3.0)	12.7 (2.7)	.16*
Word Recall	12.1 (3.0)	12.1 (3.3)	12.6 (3.1)	.06

Note. Values in parentheses are standard deviations. Health was a self-rating on a scale ranging from 1 (*excellent*) to 5 (*poor*). Scaled scores are age-adjusted scores relative to the nationally representative normative sample in the Wechsler Adult Intelligence Scale III (Wechsler, 1997a) and Wechsler Memory Scale III (Wechsler, 1997b), which have means of 10 and standard deviations of 3 in the normative sample. * $p < .01$.

may be functioning at a somewhat higher level relative to their age peers in these respects than the younger adults.

Procedure

The tasks with the word and faces stimuli were administered in the second session of the project, and the task with dot pattern stimuli was administered in the third session. Participants were instructed at the beginning of each task to pay attention to the items because their memory for that information would be tested immediately after the last item was presented.

The stimuli for the word task were identical to those used in Experiment 1. The dot pattern stimuli were created by designating seven-dot prototype patterns in a 10-by-10 grid and then creating perturbations of each prototype by varying the location of each dot in the pattern by a random amount. The bottom left panel of Figure 1 portrays a prototype and four perturbations created in this manner.

The stimuli in the faces task were created by morphing original faces 50% to a target face using Abrosoft's FantaMorph program. This morphing program uses the key dots method, in which dots are placed in corresponding locations on the two faces to be morphed. The program combines the two sets of features with the weighting based on the morphing percentage (in this case, 50%). The resulting morphed faces were then used as stimuli in the study and test phases. Most of the original faces were obtained from the AR face database (Martinez & Benavente, 1998) and the MacArthur Face Stimulus Set (<http://www.macbrain.org>), with some of the faces arbitrarily designated as targets. The right panel of Figure 1 illustrates how original faces (represented by the faces with the dotted-line borders in the outer perimeter of the panel) were morphed to a target face (in the center of the panel) to create synthetic faces that each resembled the target face. (The faces in this example were not used in the study but instead are pictures of some of the research staff who worked on the project.)

Slightly different procedures were used with the three sets of materials to try to achieve similar levels of performance. As noted in Experiment 1, with the word stimuli a single list was presented that contained 10 sets of 15 related words each and a test of 80 items (40 new, 30 old, and 10 critical lures). With faces two separate lists were presented, one with female faces and one with male faces. Each list contained four sets of 10 related faces, and the test consisted of 24 items (12 new, 8 old, and 4 critical lures). Four separate lists of dot patterns were presented, with the first and fourth lists containing two sets of 10 items each and the second and third lists containing three sets of 10 items each. The recognition tests consisted of either 12 items (6 new, 4 old, and 2 critical lures) or 18 items (9 new, 6 old, and 3 critical lures).

With each type of material, stimuli in the study phase were presented for 1 s with a 1-s interval between items, and a recognition test was administered immediately after the study list. New items in the test were selected from other stimulus sets created in the same manner as those that were studied. The test series contained an equal number of old items, and the critical lure, from each set. Participants categorized test items as "old" by pressing the "Z" key or as "new" by pressing the "/" key.

RESULTS

Figure 2 contains the means and standard errors for the HR, FAN, and FACL proportions for each type of stimulus material. For purposes of comparison, the figure also contains the results from Experiment 1. It can be seen that the overall pattern with each type of stimulus material was very similar in that the FACL proportion was closer to the HR proportion than to the FAN proportion. The FACL value was greater than the FAN value for 97% of the participants with words, for 89% of the participants with faces, and for 92% of the participants with dot pattern stimuli. The FACL proportion was greater than the HR proportion for 79% of the participants with words, for 7% of the participants with faces, and for 35% of the participants with dot pattern stimuli. Thus, there was variation across the three types of stimuli in the degree to which the FACL proportion was similar to the HR proportion, but in each case it was much greater than the FAN proportion.

The mean proportions of “old” responses and the decision times for each type of material are presented in Table 5. Inspection of the table reveals that for each type of material the FAN responses were slower than the FACL and HR responses but that the response times for HR and FACL responses were similar. As in Experiment 1, the correlations

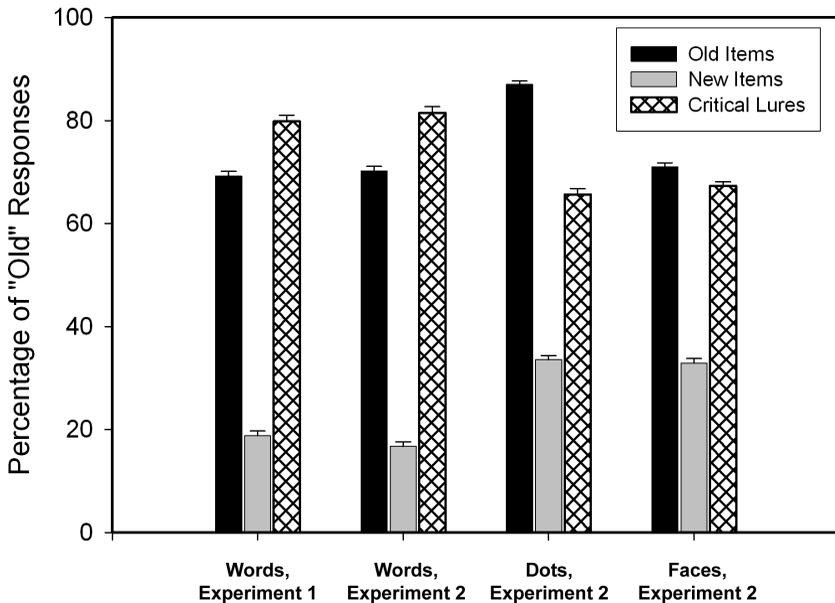


Figure 2. Means and standard errors of the proportions of “old” responses to three different types of items for the word stimuli in Experiments 1 and 2 and for the dot pattern and faces stimuli in Experiment 2

of FACL with HR were much stronger than those with FAN. People who made a large number of false recognitions therefore were likely to be higher than average in their HR proportion than in their FAN proportion. Although not portrayed in Table 5, the tendency for a higher rate of false recognitions among people with better veridical memory found in Experiment 1 was also replicated. That is, the correlations of veridical recognition (HR–FAN) and FACL were .53 for words, .23 for faces, and .34 for dots.

Four structural equation models were used to estimate the correlations between the latent constructs representing veridical recognition (HR–FAN) and false recognition (FACL) across the three tasks. One model examined correlations between the veridical memory constructs, a second examined correlations between the false recognition constructs, a third examined correlations between the residual false recognition constructs created by partialling the linear relationship of veridical memory from the FACL measures, and the fourth examined correlations between the FACL–FAN variables. In all cases, the latent constructs were represented by either two (for words and faces) or four (for dot patterns) measures derived from performance on separate parts of the tasks.

The results of these analyses are presented in Table 6. It is apparent in the first column that the model with the veridical memory (HR–FAN) measure had an excellent fit to the data, with all the variables having significant correlations with their respective latent constructs. The values in the top three rows indicate that recognition memory performance was positively correlated across the three types of materials, signifying that people who were accurate at discriminating between old and new words

Table 5. Comparisons of memory parameters, Experiment 2

	Means			Correlations with FACL	
	HR	FACL	FAN	HR	FAN
	Proportion "old"				
Words	.70 (.16)	.82 (.22)	.17 (.15)	.68*	.06
Faces	.87 (.12)	.66 (.20)	.34 (.14)	.44*	.10
Dots	.71 (.15)	.67 (.15)	.33 (.15)	.45*	-.09
	Decision time (s)				
Words	1.56 (0.75)	1.49 (0.97)	1.75 (1.20)	.85*	.59*
Faces	1.60 (0.69)	1.85 (0.86)	2.01 (0.80)	.75*	.77*
Dots	2.18 (0.90)	1.91 (0.89)	2.44 (0.97)	.83*	.77*

Note. Values in parentheses are standard deviations. FACL = false alarm to critical lures; FAN = false alarm to new items; HR = hit rate.

* $p < .01$.

also tended to be higher than average at discriminating between old and new faces and old and new dot patterns.

The models applied to the false recognition (FACL) and false recognition adjusted for veridical recognition did not fit the data as well, and several of the construct–variable correlations were not significantly different from zero. These weak correlations are a consequence of the low correlations between the variables from different parts of the task; the internal consistency (coefficient alpha) estimates of reliability were .78 for words but only .43 for faces and .22 for dot patterns. Although the false recognition constructs with the face and dot pattern stimuli are not very well defined, they nevertheless provide the best estimate of what all the variables represent because they are determined by the (reliable) variance the variables have in common.

The model with the FACL–FAN measures fit the data quite well and

Table 6. Standardized coefficients from structural models examining relationships between HR–FAN and FACL measures with different stimulus materials

	HR–FAN	FACL	FACL–FAN
Words ↔ faces	.51*	.17 (.34)	–.05
Words ↔ dots	.32*	.36 (.20)	.36*
Faces ↔ dots	.33*	.23 (.13)	.40*
Words → W1	.77*	.88* (.72*)	.79*
Words → W2	.83*	.73* (.63*)	.83*
Faces → F1	.51*	.35 (.43)	.41
Faces → F2	.68*	.76* (.58)	.67
Dots → D1	.33*	.39 (.22)	.42*
Dots → D2	.49*	.48 (.53)	.62*
Dots → D3	.80*	.23 (.17)	.51*
Dots → D4	.65*	.11 (.18)	.46*
		Fit statistics	
χ^2	26.02	37.16 (36.98)	11.46
<i>df</i>	17	18 (18)	18
CFI	.979	.914 (.828)	1.000
RMSEA	.040	.057 (.057)	.000

Note. Values in parentheses represent results based on residuals created by partialling of the influence of veridical recognition on the FACL measures. Values of the fit statistics often interpreted as representing a good fit of the model to the data are a ratio of chi-square to degrees of freedom (χ^2/df) less than 2, a confirmatory fit index (CFI) greater than .90, and a root mean squared error of approximation (RMSEA) less than .10. In order to fit models with the FACL and FACL residuals, the residual variances of the D3 and D4 variables were constrained to be equal. FACL = false alarm to critical lures; FAN = false alarm to new items; HR = hit rate. * $p < .01$.

revealed significant correlations between the measures from the tasks with words and dot stimuli and with faces and dot stimuli. However, the correlation between the corrected false recognition and gist measures with word and face stimuli was small and not significantly different from zero.

The results summarized in Table 6 imply that the tendencies to make false recognitions are only weakly correlated across different types of stimulus materials. The results portrayed in Figure 2 suggest that the false recognition phenomenon is qualitatively similar with different types of stimuli, but it may not represent a task-independent characteristic of individuals that is manifested to the same degree with different types of stimuli. This conclusion must be qualified somewhat because two of the three correlations with the corrected false recognition measure (FACL-FAN) were significant.

Table 7 lists correlations between the latent constructs representing veridical memory (HR-FAN), corrected false recognition or gist (FACL-FAN), and false recognition (FACL) with other types of variables. As in Experiment 1, the veridical memory measure with words was moderately related to episodic memory ability. However, the veridical memory measures with the other types of stimuli were not related to episodic memory ability and instead were either weakly (faces) or strongly (dot patterns) related to fluid intelligence. The strong correlation between the fluid intelligence construct and the veridical memory construct for dot pattern stimuli was not expected and may be attributable to the unfamiliar dot pattern stimuli necessitating attention to relationships between the elements, which is a strong component of fluid intelligence.

The false recognition construct for word stimuli was significantly correlated with years of education and vocabulary, but the correlations were weaker and no longer significantly different from zero after the variability in veridical memory was controlled for. These results also replicate those of Experiment 1 in the findings of no correlation of the word false recognition measure with episodic memory ability or other cognitive abilities or with personality or mood variables. The earlier results are extended by the finding of no significant correlations involving the false recognition constructs for the faces or dot pattern stimuli.

A few correlations were significant with the FACL-FAN constructs, but they were generally small and were not consistent across stimulus type. The negative age correlations for the face and dot pattern tasks were attributable largely to the positive correlation between age and FAN (i.e., .34 for faces and .21 for dots). There was also a large positive correlation between fluid intelligence and the FACL-FAN construct for dot pattern stimuli, and this reflects a combination of a negative (-.72) correlation for the FAN measure and a positive (.45) correlation for the FACL measure.

Table 7. Correlations of the memory parameters, Experiment 2

	Words			Faces			Dots		
	HR-FAN	FACL-FAN	FACL	HR-FAN	FACL-FAN	FACL	HR-FAN	FACL-FAN	FACL
Age	-.13	-.04	.02 (.10)	-.32	-.40*	-.14 (-.05)	-.34*	-.33*	-.23 (-.89)
Sex (male = 0)	.08	.04	.02 (-.02)	.10	.03	.04 (.03)	-.21*	-.21*	-.11 (-.01)
Education	.16	.18*	.17* (.13)	.05	-.12	-.10 (-.11)	.10	.12	-.00 (-.01)
Fluid intelligence	-.11	.00	-.07 (.04)	.36	.38	.04 (-.04)	.82*	.83*	.51 (.43)
Memory	.34*	.15	.03 (-.22)	.09	.05	-.03 (-.07)	-.14	-.05	.08 (.26)
Perceptual speed	.22	.16	.15 (.12)	.12	.02	.01 (.05)	.09	.01	.11 (.04)
Vocabulary	.18	.20	.22* (.17)	-.06	-.20	.00 (.02)	-.01	.04	.10 (.07)
Neuroticism	-.06	-.03	-.06 (-.04)	-.01	.06	.01 (-.04)	-.03	-.03	-.06 (-.01)
Extroversion	-.04	-.02	.04 (.12)	-.07	.04	-.01 (.06)	-.03	-.04	.05 (.02)
Openness	.13	.06	.02 (-.03)	.09	.09	.09 (.06)	.09	.06	.02 (.00)
Agreeableness	.03	-.02	-.01 (-.03)	-.01	-.01	-.03 (-.02)	-.05	-.13	-.13 (-.01)
Conscientiousness	.08	.03	.03 (-.03)	.04	.03	.03 (.06)	.00	-.07	-.10 (-.10)
Trait anxiety	-.03	-.06	-.09 (-.12)	-.01	.09	.01 (-.03)	.07	.07	-.02 (-.02)
Depression	-.06	-.08	-.07 (-.09)	-.02	.03	.01 (.06)	-.05	-.07	-.11 (-.02)
PANAS, positive	.05	.06	.04 (.04)	-.16	-.02	-.02 (.02)	-.23*	-.30*	-.13 (-.03)
PANAS, negative	-.07	-.07	-.04 (-.08)	.03	-.04	.03 (.00)	-.12	-.17	-.15 (-.11)

Note. Values in parentheses are standardized regression coefficients with residuals created by partialling of the influence of HR-FAN from the FACL measures. Several of the coefficients involving the dot pattern FACL measure before and after partialling of the influence of HR-FAN were substantial but not significant because of large standard errors. FACL = false alarm to critical lures; FAN = false alarm to new items; HR = hit rate; PANAS = Positive and Negative Affect Scale.

* $p < .01$.

Unlike in Experiment 1, the correlations between the FACL–FAN gist index and anxiety and depression were not significantly different from zero for any of the stimulus materials.

GENERAL DISCUSSION

The results of these studies replicate the false recognition phenomenon of a much greater rate of “old” responses to related new items (i.e., critical lures) than to unrelated new items and extend it to two new types of stimuli. The results also reveal that the tendency to make false recognitions can be assessed reliably at the level of individuals, particularly for the word stimuli that are most often used in research on false memories.

Both studies also found moderately high correlations of the rate of false recognitions with the proportion of old items classified as “old” (i.e., HR) but much weaker correlations with the proportion of “old” responses to unrelated new items (i.e., FAN). Therefore, not only are false recognition rates more similar at the level of means to hit rates than to false alarm rates, but the same people who tend to make many false recognitions also tend to make many correct “old” responses. This pattern is consistent with the view that with each type of stimulus material false recognitions appear to reflect processes associated with extensive encoding more than processes associated with inappropriately responding “old” to any new items. These results therefore are more compatible with an activation rather than a monitoring locus for the current false recognition results (Roediger et al., 2001).

The discovery that the false recognition rate is more strongly related to the hit rate than to the false alarm rate raises questions about the practice of subtracting or partialling the rate of false alarms to unrelated new items from the false recognition rate to adjust for the tendency to make false alarms or for a lax response criterion (Kensinger & Schacter, 1999; Sommers & Huff, 2003; Thomas & Sommers, 2005). That is, these results suggest that if the goal is to obtain an assessment of false recognition that minimizes other influences, it may be more meaningful to adjust for the tendency to correctly classify old items as “old” or for veridical memory. Nevertheless, it should be noted that the patterns of correlations for the FACL and FACL–FAN measures in Tables 3 and 7 were very similar, and thus at least in these studies the adjustment for false alarms to unrelated new items had little effect.

A particularly interesting finding was that veridical recognition of words was significantly related to episodic memory ability in both studies but that the false recognition measures were not. Although the veridical and false recognition measures had moderate correlations with one another,

largely because of the correlation between false recognition rate and hit rate, the lack of a correlation of false recognition with episodic memory ability suggests that false recognition may not be influenced by the same type of memory processing as veridical recognition. For example, false recognition may be more influenced by familiarity-based processing than veridical recognition, and the tasks used to assess the episodic memory factor are likely to involve more recollective processing than familiarity-based processing. However, the relationship between false memory and episodic memory ability may depend on the particular measures used in the assessment of each construct; Lovden (2003) found a strong negative correlation between an episodic memory construct and a false memory construct based on two recall measures and a recognition measure. Nevertheless, the current results were consistent across the two studies using the same false recognition and episodic memory variables, and thus the lack of a correlation with these particular variables appears to be robust.

The rate of false recognitions of words was also not related to other cognitive abilities, including vocabulary and fluid ability. The lack of a significant correlation with vocabulary ability suggests that it is not the case that people with more knowledge of words, and possibly richer semantic networks, are more likely to make false recognitions. People with more extensive vocabularies may have greater activation of related items, but if so, they seem able to distinguish between internal and external activation at the time of test such that they do not make any more false recognitions than people with lower levels of vocabulary knowledge.

Fluid intellectual ability measured with the same variables in this study has previously been found to have a very high correlation with a latent construct representing updating aspects of working memory and with a latent construct representing inhibition (Salthouse, Atkinson, & Berish, 2003). The absence of significant relations of fluid ability with false recognition in the current studies therefore raises the possibility that neither updating nor inhibition abilities affect the rate of false recognitions in the standard version of the DRM task. Stronger influences of fluid ability or of other cognitive abilities might be found with other versions of the task, such as recall rather than recognition (Butler et al., 2004), or when warnings about the occurrence of critical lures are presented (McCabe & Smith, 2002; Watson et al., 2005), but there is no evidence of correlations in the two independent studies reported here.

The rate of false recognition of words also was not related to various personality dimensions or to several measures of mood. Although the absence of a correlation between false recognition rates and any of the cognitive abilities might lead one to expect the tendency to make false recognitions to be related to some noncognitive factors, there was no evidence in either study that people who are more neurotic, depressed,

or anxious or report themselves to be in a negative mood make more false recognitions.

The veridical recognition measures with different types of stimuli had modest but significant correlations with one another, but the correlations between the false recognition measures were weaker and not significantly different from zero. The tendency to make false recognitions therefore appears to be at least somewhat stimulus specific and dependent on the particular type of processing carried out during the task. However, this conclusion must be considered tentative because the false recognition measures with faces and dot pattern stimuli were not very reliable, and the latent constructs based on them were not very well defined because of the weak construct–variable correlations. Furthermore, there were significant correlations for the corrected false recognition measure (FACL–FAN) between word and dot stimuli and between dot and face stimuli.

Blair et al. (2002) found good 2-week retest reliability for a false memory measure, and thus the tendency to make false memories is unlikely to be a state attribute that varies from moment to moment. One possibility that could account for the weak across-material correlations is that false recognitions may reflect how a person processes particular stimulus material. For example, rich encoding of semantic relations with words may lead to activation of semantic associates that tend to be incorrectly recognized as old, and rich encoding of physiognomic features of faces may lead to activation of physically similar features that lead to false recognitions of related faces, but the weak correlations between the false recognition rates with different stimuli suggest that it is not necessarily the same people who engage in more extensive activation with each type of stimuli.

Despite the strong negative correlation of age to the episodic memory construct in both studies (i.e., correlations of $-.44$ in Experiment 1 and $-.47$ in Experiment 2), there was no correlation between age and the tendency to falsely recognize related new items as old. There was a significant positive correlation of age to false recognitions for “remember” decisions in Experiment 1, but that could be artifactual because the age correlation was negative for “know” decisions, and the proportions of the two types of decisions are inversely related to one another. There were also no correlations of age with the difference between hit rate and false recognition rate proportions or decision times in either study. Although age differences in false recognition rate have been reported under some conditions, such as when the participant makes a forced-choice discrimination between old and critical lure items (LaVoie & Faulkner, 2000) or use warnings (McCabe & Smith, 2002), the results of these studies suggest that the false recognition phenomenon often is quantitatively similar across the period of adulthood.

Tun et al. (1998) suggested that “an age-related increase in gist-based

processing may underlie . . . age differences in false memory,” and similar views were expressed by Kensinger and Schacter (1999) and Koutstaal and Schacter (1997). However, the lack of age differences in the rate of false recognitions and the failure to find a positive correlation of age with the (FACL–FAN) gist index for words are inconsistent with this interpretation. Moreover, the correlation between age and the gist index was actually negative rather than positive for the face and dot pattern stimuli.

In summary, the results of these studies indicate that the false recognition phenomenon can be measured reliably at the level of the individual and that it is closely related to the tendency to say “old” to old items, which suggests that it is probably attributable to activation of related items at encoding. However, the rate of false recognitions was not related to episodic memory, several other cognitive abilities, personality dimensions, or mood. Because false recognition rates with words, faces, and dot pattern stimuli were not consistently correlated with one another, false recognition may reflect aspects of processing associated with a particular type of stimuli rather than a general, or task-independent, characteristic of individuals.

Notes

This research was supported by National Institute on Aging Grant 19627, awarded to Timothy Salthouse. We would like to thank the following people for assistance in data collection, scoring, and data entry: Irina Bocarnea, Kristina Caudle, Edison Choe, Steven Cholewiak, Elise Clerkin, Maggie Davis, Jing Fang, Crystal Gomez, Paul Hiatt, Lindsey Jones, Katherine Kane, Carolyn Kilday, Josh Magee, Elliott Neal, Nicole Numbers, Laura Orbann, Cris Rabaglia, Erycka Reid, Jesse Schneider, Leigh Schoettinger, Julia Siegel, Mary Thibadeau, Elliot Tucker-Drob, and Laura Wells.

Correspondence about this article should be addressed to Timothy A. Salthouse, Department of Psychology, 102 Gilmer Hall, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904-4400 (e-mail: Salthouse@virginia.edu). Received for publication January 2, 2006; revision received April 1, 2006.

References

- Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L. III, McDermott, K. B., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology*, *16*, 361–384.
- Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 941–947.
- Blair, I. V., Lenton, A. P., & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review*, *9*, 590–596.
- Budson, A. E., Daffner, K. R., Desikan, R., & Schacter, D. L. (2000). When false

Appendix. Standardized regression coefficients for confirmatory factor analyses conducted on the reference cognitive ability variables, Experiments 1 and 2

	Fluid ability	Episodic memory	Perceptual speed	Vocabulary
Raven's	.84, .85			
Letter Sets	.71, .72			
Shipley Abstraction	.80, .82			
Spatial Relations	.82, .80			
Paper Folding	.79, .81			
Form Boards	.71, .77			
Word Recall		.74, .78		
Logical Memory		.65, .65		
Paired Associates		.77, .77		
Digit Symbol			.81, .70	
Letter Comparison			.80, .66	
Pattern Comparison			.85, .85	
Wechsler Adult Intelligence Scale				
III Vocabulary				.88, .88
Woodcock Johnson Picture Vocabulary				.81, .82
Synonym Vocabulary				.87, .89
Antonym Vocabulary				.82, .83

Note. The first number is the coefficient from Experiment 1, and the second is the coefficient from Experiment 2.

Correlations between cognitive ability constructs

	Fluid ability	Episodic memory	Perceptual speed	Vocabulary
Fluid ability	X	.64	.67	.52
Episodic memory	.67	X	.56	.39
Perceptual speed	.66	.49	X	.22
Vocabulary	.48	.38	.07	X

Note. Entries above the diagonal are correlations from Experiment 1, and those below the diagonal are from Experiment 2.

recognition is unopposed by true recognition: Gist-based memory distortion in Alzheimer's disease. *Neuropsychology*, 14, 277-287.

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L. III. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review*, 11, 921-925.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58, 17-22.

- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.
- Gallo, D. A., & Roediger, H. L. III. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition, 31*, 1036–1044.
- Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18*, 579–583.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, the Netherlands: Tilburg University Press.
- Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1999). Age, testing at preferred or nonpreferred times (testing optimality) and false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 23–40.
- Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology, 16*, 399–415.
- Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 499–510.
- Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*, 555–583.
- Koutstaal, W., Schacter, D. L., & Brenner, C. (2001). Dual task demands and gist-based false recognition of pictures in younger and older adults. *Journal of Memory and Language, 44*, 399–426.
- Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14*, 220–237.
- LaVoie, D. J., & Faulkner, K. (2000). Age differences in false recognition using a forced choice paradigm. *Experimental Aging Research, 26*, 367–381.
- Lovden, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*, 268–283.
- Martinez, A. M., & Benavente, R. (1998). *The AR face database*. CVC Technical Report no. 24. Available at http://rv11.ecn.purdue.edu/~aleix/aleix_face_DB.html.
- McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition, 30*, 1065–1077.
- Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304–308.

- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim thesis. *Learning and Individual Differences, 7*, 1–75.
- Roediger, H. L. III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.
- Roediger, H. L. III, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407.
- Salthouse, T. A. (2004). Localizing age-related individual differences in a hierarchical structure. *Intelligence, 32*, 541–561.
- Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*, 532–545.
- Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566–594.
- Salthouse, T. A., Berish, D. E., & Siedlecki, K. L. (2004). Construct validity and age sensitivity of prospective memory. *Memory & Cognition, 32*, 1133–1148.
- Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24.
- Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology, 72*, 499–503.
- Sommers, M. S., & Huff, L. M. (2003). The effects of age and dementia of the Alzheimer's type on phonological false memories. *Psychology and Aging, 18*, 791–806.
- Spielberger, C. D., Gorsuch, R. L., & Lushere, R. E. (1970). *Manual for the State-Trait Anxiety Inventory: Form X*. Palo Alto, CA: Consulting Psychologists Press.
- Stadler, M. A., Roediger, H. L. III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition, 27*, 494–500.
- Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy, with happiness false memory: Mood and the false memory effect. *Psychological Science, 16*, 785–791.
- Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*, 71–86.
- Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*, 230–241.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.
- Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology, 15*, 254–267.

- Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese–Roediger–McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 76–85.
- Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler Memory Scale* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- Winograd, E., Peluso, J. P., & Glover, T. A. (1998). Individual differences in susceptibility to memory illusions. *Applied Cognitive Psychology*, *12*, S5–S27.
- Wright, D. B., Startup, H. M., & Mathews, S. A. (2005). Mood, dissociation and false memories using the Deese–Roediger–McDermott procedure. *British Journal of Psychology*, *96*, 283–293.