When false recognition is out of control: The case of facial conjunctions

Todd C. Jones
Victoria University of Wellington, Wellington, New Zealand

and

James C. Bartlett
University of Texas at Dallas, Richardson, Texas

In three experiments, a dual-process approach to face recognition memory is examined, with a specific focus on the idea that a recollection process can be used to retrieve configural information of a studied face. Subjects could avoid, with confidence, a recognition error to conjunction lure faces (each a reconfiguration of features from separate studied faces) or feature lure faces (each based on a set of old features and a set of new features) by recalling a studied configuration. In Experiment 1, study repetition (one vs. eight presentations) was manipulated, and in Experiments 2 and 3, retention interval over a short number of trials (0–20) was manipulated. Different measures converged on the conclusion that subjects were unable to use a recollection process to retrieve configural information in an effort to temper recognition errors for conjunction or feature lure faces. A single process, familiarity, appears to be the sole process underlying recognition of conjunction and feature faces, and familiarity contributes, perhaps in whole, to discrimination of old from conjunction faces.

Many conceptions of face processing hold that both individual features and configural information support face recognition (e.g., Cabeza & Kato, 2000; Carey & Diamond, 1977; Farah, Wilson, Drain, & Tanaka, 1998; Rakover, 2002; Searcy & Bartlett, 1996). One finding that supports this conceptualization comes from the feature–conjunction paradigm (e.g., T. C. Jones, Bartlett, & Wade, 2006; Reinitz, Lammers, & Cochran, 1992; Reinitz, Morrissey, & Demb, 1994; Searcy, Bartlett, & Memon, 1999). In this paradigm, critical recognition test lures contain either one set of studied features and one set of new features (called *feature lures*) or sets of studied features reconfigured across separate study faces (called *conjunction lures*—e.g., Reinitz et al., 1992; or *feature prototypes*—Cabeza & Kato, 2000; see examples in Figure 1). False alarm rates to feature and conjunction lures (referred to as *feature effects* and *conjunction effects*, respectively) are typically above those for (wholly) new faces, and these effects indicate the influence of featural information. Typically, conjunction effects are significantly larger than feature effects (e.g., T. C. Jones et al., 2006), owing to more studied features being present in conjunction faces than in feature faces (i.e., a greater match or overlap with studied features). The hit rate for studied faces is nearly always higher than is the false alarm rate for conjunction faces, indicating an influence of configural information. For both old and conjunction faces, all of the features are old. What differs is the configuration of these features.[1]

Another finding from the conjunction paradigm that supports the feature-configuration distinction comes from studies that have used an inversion manipulation, which has been thought to disrupt the processing of configural information much more than it disrupts featural information (e.g., Searcy & Bartlett, 1996; but see Rakover & Teucher, 1997). The idea that face inversion disrupts the processing of configural information leads to the predictions that old–conjunction discrimination should be lower for inverted than for upright faces and that conjunction–new discrimination (i.e., the conjunction effect) should be largely unaffected. Indeed, these predictions have been borne out (e.g., Bartlett, Searcy, & Abdi, 2003; Cabeza & Kato, 2000; McKone & Peh, 2006).

Although many researchers have focused on the contributions from features and configurations to face recognition memory, we question how featural and configural information might be retrieved. For example, a single-process (i.e., familiarity strength) model (e.g., Banks, 1970) that is based on the amount of perceptual overlap of a test lure and earlier study faces could account for the pattern of "old" responses described above (i.e., hits > conjunction lures > feature lures > new; T. C. Jones et al., 2006). However, some researchers have adopted a general dual-process framework (e.g., Mandler, 1980; for reviews, see Diana, Reder, Arndt, & Park, 2006; Yonelinas, 2002) where familiarity and a slower, controlled process, recollection, contribute to face recognition memory

T. C. Jones, todd.jones@vuw.ac.nz
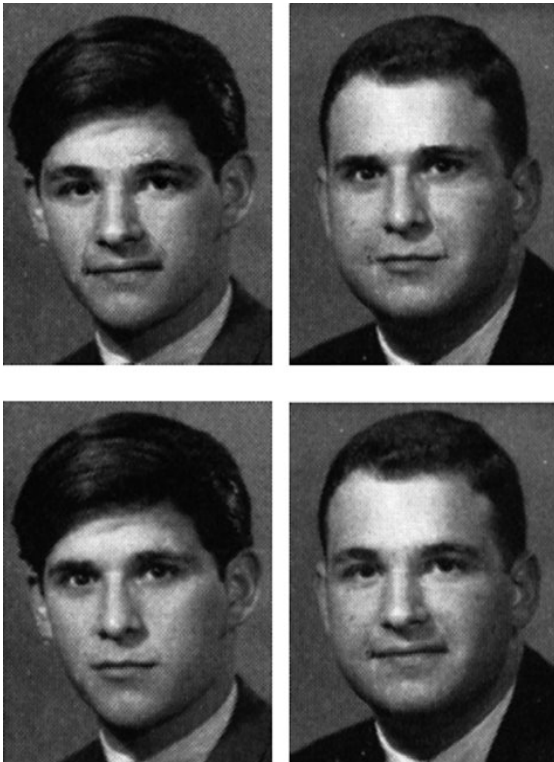
Figure 1. Example stimuli. The top two faces are constructions created by swapping the inner features (eyes, eyebrows, nose, and mouth) from the bottom two original faces (*The Hawkeye*, 1963).

(Bartlett et al., 2003; Bastin & Van der Linden, 2003; Mäntylä, 1997; Mäntylä & Cornoldi, 2002; Mäntylä & Holm, 2005). The present research examined the viability of a dual-process approach, where, to avoid conjunction errors, recollection can be used to retrieve configural information.[2]

One line of evidence that supports a role of recollection for configural information comes from applications of the remember–know procedure (Gardiner, 1988; Tulving, 1985). For example, one of these studies used an inversion manipulation at study and test (in a 2 × 2 design) to impair the encoding or retrieval of configurations (Mäntylä & Holm, 2005). The stimuli were human faces and horse faces. Upright human faces elicited a much higher rate of remember responses in comparison with the other (seven) conditions, but all of the conditions produced an equivalent rate of know judgments. The authors concluded that configural information for human faces supports recollection-based recognition, as measured by remember judgments, but not familiarity-based recognition, as measured by know judgments. In another study, the rate of remember responses was higher following two spaced study presentations than following two massed study presentations (Mäntylä & Cornoldi, 2002). Together, these studies suggest that recollection of configural information occurs in upright human face recognition and is increased by spaced repetitions.

However, other findings from remember–know studies with faces question the conclusion that configurations support facial recollection. First, using face drawings, Reinitz et al. (1994) found that conjunction and feature lure faces—both of which contained nonstudied configurations—were judged "remember" at rates above that for a new face baseline (e.g., .35 and .16, respectively, as opposed to .02 for a full attention condition). Similarly, after presenting study faces twice, Bartlett et al. (2003) found an elevated remember rate for naturalistic conjunction faces (.11) compared with that for new faces (.04). What these results suggest is that configural information may not support remember judgments exclusively. In addition, other researchers have provided evidence that remember judgments may not provide a clean measure of recollection, because remember judgments may sometimes be based on high familiarity (Dougal & Rotello, 2007; Rotello, Macmillan, Reeder, & Wong, 2005; see also Donaldson, 1996; Dunn, 2004, 2008; Xu & Malmberg, 2007). Thus, there should be at least some concern for how well remember judgments measure recollection in faces, including facial configurations.

A second line of research that suggests that recollection may be involved in the facial conjunction paradigm comes from research on compound word conjunctions (e.g., study, *raindrop* and *crossbow*; test lure, *rainbow*). This research using compound word conjunctions demonstrates nothing about the role of recollection in the domain of faces. However, the methods in those studies have provided evidence for the use of recollection and allow for some clear tests of a single- versus a dual-process approach for face recognition memory. Thus, these studies are relevant to the present aims.

Some of these studies with compound word materials have manipulated study repetition (T. C. Jones, 2005; T. C. Jones & Jacoby, 2001; Lampinen, Odegard, & Neuschatz, 2004), which is thought to increase familiarity and recollection (Yonelinas, 2002). Some studies (T. C. Jones, 2005; T. C. Jones & Jacoby, 2001) have relied on the idea that, under a short response deadline, responses are based heavily on familiarity, whereas, at a long response deadline, responses can be made by using a slower, controlled recollection process (e.g., Jacoby, T. C. Jones, & Dolan, 1998; McElree, Dolan, & Jacoby, 1999; Rotello, Macmillan, & Van Tassel, 2000). Study repetition significantly increased feature and conjunction effects when time to use recollection was limited (T. C. Jones, 2005; T. C. Jones & Jacoby, 2001), showing that study repetition increased familiarity. However, study repetition did not affect feature and conjunction effects significantly when more time to use recollection was given (T. C. Jones, 2005; T. C. Jones & Jacoby, 2001; see also Lamipinen et al., 2004), indicating that study repetition also increased recollection. Under the long response deadline condition, an increase in recollection (i.e., with a recall-to-reject strategy) was thought to offset the increase in familiarity (see also Jacoby et al., 1998; Kelley & Wixted, 2001; Light, Patterson, Chung, & Healy, 2004).
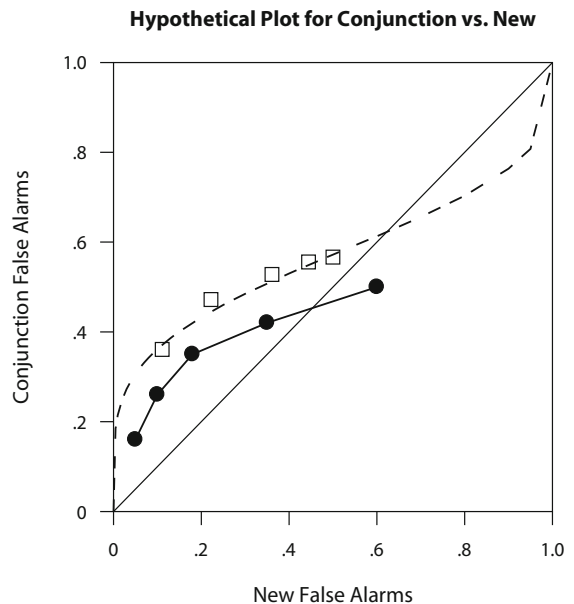
Support for recollection-based rejections was particularly strong in the study by Lampinen et al. (2004). They obtained confidence ratings on all test trials and

later solicited the reasons for correct rejections of conjunction and feature test lures: (1) Study repetition increased the likelihood that a feature or conjunction lure was rejected with strong confidence. (2) These confident correct rejections were often backed up by recall of an actual parent word, clearly demonstrating a successful recall-to-reject strategy. (3) Lampinen et al. used receiver-operating characteristic (ROC) curves (or memory operating characteristic, MOC, functions; see, e.g., Banks, 1970) to identify rejection of conjunctions with a recollection process.

In this last approach, instead of plotting the hit against false alarm rate at cumulative levels of confidence, as is typical for ROC functions, one plots the false alarm rate of conjunctions against the false alarm rate of new items (see Kelley & Wixted, 2001).[3] ROC functions usually do not cross the unity diagonal, but an empirical point could be below chance, producing a dip in the empirical ROC function below the unity diagonal (see Figure 2). Alternatively, none of the empirical points might fall below chance, but the fitted ROC function could still dip across the unity diagonal (see Figure 2). A *dip effect* (Heathcote, Raymond, & Dunn, 2006) can reflect a greater proportion of strong-confidence correct rejections for critical lures than for new, baseline items, and a dip effect has been interpreted as evidence for the use of recollection in rejecting lures (e.g., Heathcote et al., 2006; Lampinen et al., 2004; see also Kelley & Wixted, 2001). The key point is that Lampinen et al. (2004) found a dip effect for conjunction–new discrimination, particularly when the parent words were presented four times in the study phase.

In another set of verbal studies, the retention interval (lag, 0–20 trials) was manipulated in a single, long series of recognition trials, where compound words were presented one at a time and were later repeated (i.e., a continuous recognition memory task; Shepard & Teghtsoonian, 1961). T. C. Jones and Atchley (2002, 2006) found that, although hit rates to old words decreased across lag retention intervals, a quadratic trend occurred for the conjunction error rates: The error was relatively low at a 0-trial lag; it increased at a 1-trial lag and then gradually decreased out to a 20-trial lag (see Figure 3A). The relatively low error rate at the 0-trial lag indicated a relative increase in the ability to use a controlled process at the very short lag. Moreover, in the later study (T. C. Jones & Atchley, 2006), subjective judgments of parent word retrieval and actual recall of parent words showed that recollection-based rejections of conjunction lures occurred across all lags; the rate of this recollection-based rejection was robust at the 0-trial lag; it then dropped dramatically and stayed relatively constant from the 1-trial lag out to the 20-trial lag (see Figure 3B). Thus, the findings indicated that the rate of recollection-based rejections of the conjunctions was robust at lag 0 but was much reduced as the lag increased above lag 0.

According to the single-process (*familiarity strength*) approach, conjunction faces in a strong memory condition (multiple study presentation or shorter retention interval) should be more familiar than conjunction faces in a weak memory condition (single study presentation or longer

Hypothetical Plot for Conjunction vs. New

Figure 2. Hypothetical conjunction–new ROC functions giving evidence of recollection-based rejections of conjunctions. The solid line with circle symbols shows an ROC function with an empirical point below the unity diagonal; the empty squares show empirical points all above the unity diagonal with a fitted ROC function that crosses the unity diagonal. Both of these functions show a dip effect.

retention interval). Thus, this approach predicts that there should be (1) a higher conjunction error rate for strong memory conditions than for weak memory conditions and (2) more confident correct rejections in weak memory conditions than in strong memory conditions. This approach also predicts a conjunction–new ROC function that does not cross the unity diagonal.

By contrast, the dual-process approach allows for the possibility that configural information can be recalled and used to avoid false recognition errors. We refer to this possibility as the *configuration–recollection hypothesis*, and, in the present experiments, we test two predictions that it makes for the facial conjunction paradigm. The first prediction is that conjunction effects should be lower, or at least not increased, for conditions (e.g., multiple study repetitions, short retention interval) that should increase the probability of using recollection successfully. The second prediction, which might save the configuration–recollection hypothesis if the first prediction fails, comes from a counterintuitive finding emerging recently in the literature: A condition that produces a relatively high compound-word conjunction error rate can also produce a relatively high subjective (i.e., judged) parent recall rate (Arndt & T. C. Jones, 2008) or objective (i.e., actual) parent recall rate (Arndt & T. C. Jones, 2008; Odegard, Lampinen, & Toglia, 2005). Thus, regardless of whether one observes an increase in the error rate for the strong memory conditions, one might still observe an increase in confident correct rejections for strong memory
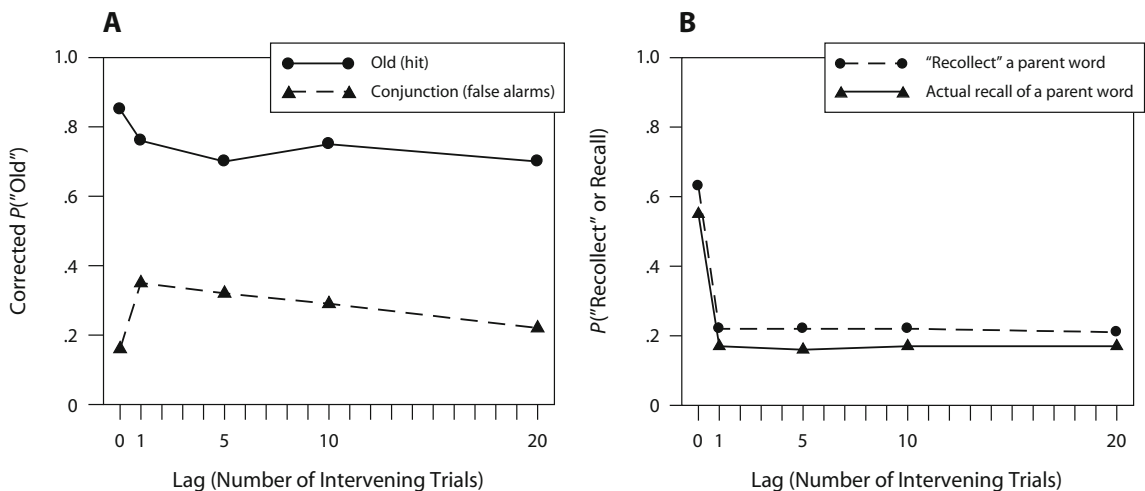
Figure 3. (A) "Old" responses for old (hits) and conjunction items (false alarms), corrected for baseline false alarms (.04 for Experiment 2A and .05 for Experiment 2B) by simple subtraction of the baseline (the lag 5 is a mean from the two experiments) from T. C. Jones and Atchley's (2006) experiments. **(B)** Measures of recollection-based rejection of conjunction lures from T. C. Jones and Atchley's (2006) Experiments 2A and 2B (the lag 5 is a mean from the two experiments); false recollect or recall responses for new words were rare (.02 and .00, respectively).

conditions relative to weak memory conditions (e.g., Xu & Malmberg, 2007), as well as a conjunction–new ROC function that dips across the unity diagonal. If there are more confident correct rejections in the strong memory conditions, and if a shift in bias can be ruled out, the findings would support the configuration–recollection hypothesis.

In addition to examining the predictions above, we took the extra measure of manipulating the test instructions in Experiments 1 and 2. We used test instructions that explicitly directed subjects to treat conjunction faces as lures, and we used test instructions that directly (or tacitly) encouraged subjects to treat the conjunction faces as targets. These two instruction types can be viewed as corresponding to exclusion and inclusion instructions used with the process dissociation procedure (Jacoby, 1991).

A single-process approach predicts very similar results for the two test instruction conditions, with higher "old" response rates for old and conjunction faces in relatively strong memory conditions (multiple study presentations, shorter retention intervals) than in weak memory conditions (a single study presentation, longer retention interval). However, if subjects can control face conjunction errors with a recollection process, the responses to conjunction faces should produce a different pattern, depending on the two test instruction conditions. For example, an inclusion (or no warning) instruction should produce higher "old" response rates for conjunction faces in the strong memory conditions than in the weak memory conditions. In contrast, an exclusion (or warning) instruction should produce the opposite pattern or no change in conjunction effects, as has been found in the verbal domain. Such outcomes would provide support for a dual-process approach to unknown face recognition.

EXPERIMENT 1

In this experiment, we examined the effect of study repetition on conjunction memory errors for naturalistic faces of unknown individuals. As noted above, Mäntylä and Cornoldi (2002) concluded that spaced repetition of study faces led to greater recollection than did massed repetition of study faces, and Mäntylä and Holm (2005) concluded that configural information fosters recollection. These two conclusions would lead one to suspect that repeated study repetitions would afford a greater opportunity for recollection to be used to avoid a conjunction error.

On the other hand, Bartlett et al. (2003) described two experiments that examined study repetition with conjunction faces, where the "parent" (study) faces were presented one or three times (between subjects) or one or eight times (within subjects). Study repetition increased the hit rate for old faces, as one would expect, but, ironically, increased the conjunction effect. (The same pattern occurred for inverted faces, despite a sharp reduction in overall accuracy of performance.) The experiments described by Bartlett et al. indicated that study repetition increased familiarity, which increased conjunction errors. However, they did not rule out the possibility that recollection was used to lower conjunction errors, at least to some extent.

We presented study faces one or eight times in our present experiment, and our procedures allowed for an assessment through the analysis of error rates, confident correct rejections, and ROC functions in two test instruction conditions. All subjects took an exclusion test and an inclusion test. For each set of instructions, the confidence rating scale was explained, and subjects were told that some faces would be exact matches to studied faces, that some of the faces would be recombinations (for which

parts of separate studied faces had been recombined), and that some of the faces would be entirely new (for which none of the parts had been presented earlier). Three example test stimuli (one exact match, one recombination, and one entirely new face) were shown before each test. Each face was presented by itself and was preceded by "This following is an example of a(n) _____ face," where "old," "recombination," or "new" appeared in the blank. For the exclusion instructions, the subjects were told that the correct response for exact matches to study faces was "old," but that the correct response for conjunction and new faces was "new." For the inclusion instructions, the subjects were told that the correct response for exact matches and recombination faces was "old," but that the correct response for new faces was "new."

We expected that study repetition would increase "old" responses for old faces similarly for the two sets of test instructions. We also expected that, for the inclusion test condition, study repetition would increase "old" responses for conjunction faces. However, for the exclusion test condition, the effect of study repetition on conjunctions might be smaller, ameliorated, or reversed. Such a finding for the exclusion condition would support the configuration–recollection hypothesis. This hypothesis would also be supported if study repetition increased high-confidence rejections of conjunction lures (cf. Lampinen et al., 2004). Finally, we used the inclusion and exclusion confidence data to examine ROC functions for discrimination of conjunction from new faces. Again, an ROC function that crosses the unity diagonal would be taken as an indicator of successful rejection of lures on the basis of recollection. This dip effect should occur for the exclusion condition but not for the inclusion condition, with several study repetitions being more likely to produce a dip effect.

Method

Subjects. Sixty Victoria University of Wellington undergraduates participated and earned credit toward an introductory psychology research requirement.

Apparatus. The stimuli were the grayscale face photographs of young men used by T. C. Jones et al. (2006). These stimuli show the whole head and are cropped across the shoulders. Each subject of the photographs wore a dark suit, light shirt, and a tie. None of the subjects of the photographs sported facial hair or spectacles. The original photographs appeared in *The Hawkeye* (1963), and constructions were made (in Adobe Photoshop) by exchanging inner features (eyes, nose, mouth) of two faces with a similar pose (see Figure 1). In the present experiments, constructed faces served as "parents" for the conjunction condition and the originals served as old, conjunction, or new faces on the tests. There were 100 pairs (50 original pairs, 50 constructed pairs) of critical items. An additional 34 faces were used as primacy or recency buffers in the study phase and as test examples. The experiments were run on PCs with an E-Prime program (Schneider, Eschman, & Zuccolotto, 2002a, 2002b) and 15-in. monitors. Up to 5 subjects participated at a time at individual computer stations.

To establish that subjects could discriminate original faces from created faces containing the inner or outer features of an original face, we first ran an experiment ($n = 16$) in which two faces were presented simultaneously for 4 sec on each trial. The faces were either the same (side-by-side copies of an original) or different (e.g., an original beside a construction containing the inner or outer features of the adjacent original), and subjects made either a "same" or

a "different" response. The mean accuracy rates were .95 for same-face pairs and .93 for different-face pairs. Thus, in a simultaneous viewing procedure, subjects could readily identify the differences between two faces sharing either the inner or the outer features.

Design. All factors were manipulated within subjects. There were two tests with separate test instructions, one for inclusion and one for exclusion. There were five types of items on each test: old–presented one time at study (old–1P), old–presented eight times at study (old–8P), conjunction–parents presented one time at study (conjunction–1P), conjunction–parents presented eight times at study (conjunction–8P), and new. Nested within the item type factor was a 2 (item type: old, conjunction) $\times$ 2 (number of study presentations: one vs. eight) design.

Procedure. Two sets of faces (10 pairs of originals, 10 pairs of parents) were presented in seven blocks of trials. There were 44 trials (2 primacy buffers, 40 critical stimuli, 2 recency buffers) in these first seven blocks. Each face was shown (alone) once per block for 2,000 msec in the center of the screen with a black background. The ITI was 250 msec. The order of the faces changed from block to block, and the two parent faces (always constructions) for later conjunction faces were separated by 2–13 trials (mean parent–parent lag = 7.3 trials). In an eighth block of trials, there were 84 trials (2 primacy buffers, 80 critical stimuli, 2 recency buffers). The faces from Blocks 1–7 were presented again but mixed among two additional sets of faces (10 pairs of originals, 10 pairs of parents) for the single presentation conditions. The faces for the different study conditions were distributed throughout the list, and for both repeated parent faces and single-presentation parent faces, the parent–parent lag ranged from 2 to 6 trials ($M = 4$ trials).

Two tests, each with its own set of instructions, followed the study phase. There were 50 faces on each test: 10 old–1P, 10 old–8P, 10 conjunction–1P, 10 conjunction–8P, and 10 new. The test faces were rotated through the 10 different cells (2 test types $\times$ 5 study conditions), so that each face appeared in each cell equally often. The order of the test types was also counterbalanced across subjects. The 5 study conditions were distributed evenly throughout each test. Before each test, the use of a 6-point rating scale was explained and 3 example test faces (1 old, 1 conjunction, 1 new) were presented. Each example face appeared on the screen for 8 msec.

In the (conjunction) exclusion test instructions, subjects were directed to identify as "old" only exact matches. More specifically, they were told, "For this test, you should judge a face as 'old' if it appears exactly as it did earlier. You should judge a face 'new' if the parts in the face are old but the face, as a whole, is new. That is, you should judge recombination faces—ones that are similar (but different) from faces seen earlier—as 'new.'" In the inclusion test instructions, subjects were directed to identify as "old" both exact matches and recombination faces. Again, specifically, they were told, "For this test, you should judge a face as 'old' if it appears exactly as it did earlier. Also, you should judge a face 'old' if the parts in the face are old but the face, as a whole, is new. That is, you should judge recombination faces—ones that are similar (but different) from faces seen earlier—as 'old.'" Subjects were encouraged to ask questions if they felt unclear about the test instructions.

On each test trial, a face appeared alone for an initial 500 msec, during which time a response was not accepted by the computer. A rating scale (1 = *very sure new*, 2 = *somewhat sure new*, 3 = *unsure new*, 4 = *unsure old*, 5 = *somewhat sure old*, 6 = *very sure old*) then appeared below a test face, with the test face on the screen. From the onset of the presentation of the rating scale, subjects were given up to 8 sec to enter a response. If no response was entered, the program continued to the next trial. The ITI was 500 msec. For each experiment, a test face appeared on only one test.

Results and Discussion

The proportions of "old" responses (i.e., 4–6) are shown in Table 1 for each test condition and item type (with the total number of trials within a condition as the denomi-

Table 1
Mean Proportion of "Old" Judgments by Test and Item Type

	Old				Conjunction				New	
	1P		8P		1P		8P			
Test Type	M	SD	M	SD	M	SD	M	SD	M	SD
Inclusion	.51	.20	.90	.12	.48	.19	.73	.19	.43	.19
Exclusion	.44	.21	.82	.18	.38	.18	.61	.22	.31	.19

nator). As can be seen, the two sets of test instructions produced the same pattern of data, with the exclusion test producing "old" responses consistently below those for the inclusion test. The baseline false alarm rates for new faces was lower on the exclusion test than on the inclusion test [$F(1,59) = 11.43$, $MS_e = 0.04$]. (For all analyses, $\alpha = .05$.) Thus, the effect of test instructions simply appeared to shift subjects' willingness to respond "old."

The proportion of "old" responses for old faces was higher than that for conjunction faces, which was higher than that for new faces. For the exclusion test, the conjunction error rates for the two study presentation conditions were higher than the baseline error rate, demonstrating the standard conjunction effect. Study repetition increased the proportion of "old" responses for both old faces and conjunction faces, and these increases were very similar for the two test conditions. A 2 (test instruction: inclusion, exclusion) $\times$ 2 (item type: old, conjunction) $\times$ 2 (number of study presentations: one, eight) repeated measures ANOVA on the "old" response scores with the baseline error rate subtracted (i.e., on corrected recognition) produced significant main effects of test instruction [$F(1,59) = 15.59$, $MS_e = 0.07$], item type [$F(1,59) = 96.02$, $MS_e = 0.02$], and study repetition [$F(1,59) = 401.79$, $MS_e = 0.03$]. The test instruction effect was not qualified by any higher order interactions, meaning that study repetition and item type had similar effects in the two test instruction conditions. The only significant interaction occurred between item type and study repetition [$F(1,59) = 21.16$, $MS_e = 0.03$], with study repetition increasing "old" responses more for old faces than for conjunction faces (all other Fs < 1.63). Therefore, there was no evidence from these data that subjects were able to recall the configuration of a parent face to reject a conjunction lure.

We next compared "very sure new" rates on the exclusion test for the conjunction faces whose features had been studied one or eight times. Again, the configuration–recollection hypothesis would be supported by an increase in "very sure new" responses with additional study presentations. The mean proportion of "very sure new" responses *decreased* significantly with additional study presentations (conjunction–1P = 0.25, conjunction–8P = 0.15) [$F(1,59) = 14.71$, $MS_e = 0.01$]. (The same pattern occurred for the inclusion test.) Thus, by this measure, there was no evidence that study repetition increased control of the face conjunction errors.

Finally, we used J. Eng's (n.d.) program based on the standard signal-detection model of Green and Swets (1974) to produce conjunction–new ROC functions. For this approach, we used the confidence level frequencies

of the group because (1) the ROC functions could not be obtained for a small number of individual subjects (cf. Kelley & Wixted, 2001), (2) we were concerned for the stability of the ROC functions at the subject level, and (3) this was the approach used by Lampinen et al. (2004). For our purposes, no formal statistics on any specific model parameters were required (or desired).[4] Our aim was to assess whether any of the empirical conjunction–new points or fitted conjunction–new ROC functions for the exclusion test dipped across the unity diagonal, and we compared (informally) the ROC functions for the two test instruction groups. As can be seen in Figure 4, none of the empirical points and fitted ROC functions dipped across the unity diagonal. Perhaps more impressive is the marked similarity in the ROC functions for the exclusion and inclusion tests, both of which showed greater conjunction–new discrimination in a strong memory condition (eight study presentations) than in a weak memory condition (one presentation). This outcome offers no support for the configuration–recollection hypothesis, indicating that subjects were unable to control conjunction errors with a recollection process.
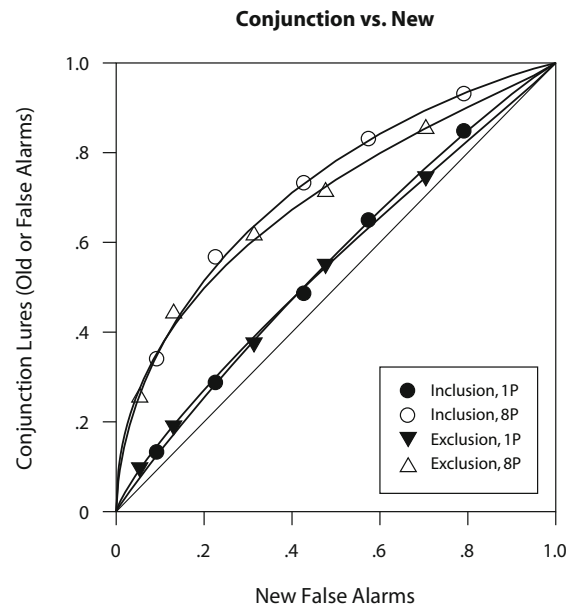
Conjunction vs. New

Figure 4. Experiment 1: Conjunction–new ROC functions for conjunction-as-old and conjunction-as-lure conditions and study repetition conditions. 1P = one presentation during study; 8P = eight presentations during study.

To summarize, the "old" response rates, confident correct rejections, and ROC functions converge on the conclusion that subjects were unable to use recollection to retrieve an earlier configuration, even when the earlier configurations had been presented eight times.

EXPERIMENT 2

In T. C. Jones and Atchley's (2006) continuous recognition memory experiments on compound word conjunction errors, recollection-based rejections of conjunction lures occurred at lag retention intervals from 0 to 20 trials. We adopted a continuous recognition test to investigate whether conjunction errors to faces could be controlled at these intervals by employing the tactics used in Experiment 1 (i.e., test instruction manipulation [forewarning of conjunction lures or no forewarning of conjunction lures] and confidence judgments). With respect to conjunction error rates, one might expect a quadratic trend similar to that obtained by T. C. Jones and Atchley (2002, 2006), but this pattern would not indicate whether a recall-to-reject strategy was successful beyond the 0-lag retention interval. To ascertain whether a recall-to-reject strategy was successful beyond the 0-trial lag, we turned to confident correct rejections and ROC functions. If recollection-based rejections occur in this paradigm, there should be more confident correct rejections for the earlier lags, particularly the 0-trial lag, than for the later lags. Also, if recall of earlier faces to reject conjunction lures occurs at lags from 1 to 20 trials, the conjunction–new ROC functions for these lag conditions should all show a dip effect. If the rate of image recall is relatively constant from 1 to 20 trials (cf. T. C. Jones & Atchley, 2006), the ROC functions should be similar to each other but should still show a dip effect.

We again compared performance for two sets of test instructions. One group of subjects was explicitly forewarned about the presence of conjunction lures and was told how they could avoid an error by recalling how an earlier face looked (exclusion instructions). A second group of subjects was not warned about the conjunction lures. (Anecdotal feedback from subjects for other face conjunction experiments [e.g., T. C. Jones et al., 2006] indicated that the lack of a warning about the face conjunction lures amounted to a tacit inclusion instruction, but, for the sake of precision, we refer to the conditions as *warned* and *not warned*.)

Method

Subjects. Victoria University of Wellington undergraduates ($N = 160$) participated voluntarily or for credit toward a class research requirement. The subjects were assigned randomly to one of two groups ($n = 80$ per group). The subjects sat at individual, partitioned stations, and as many as 16 subjects were tested in a given session.

Apparatus. A subset (80 critical, 80 sets of parents, and 20 fillers) of the faces from Experiment 1 were used, and a computer program created with E-Prime software (Schneider et al., 2002a, 2002b) was run on individual PCs.

Design. There were two groups of subjects. One received instructions warning about conjunction lures; the other did not. The type of item (old or conjunction) and the retention interval (lag in number of trials) were manipulated within subjects. The first presentations of faces that were shown again later (as old faces) served as new faces.

Procedure. A series of 200 faces was presented, and, within this series, we manipulated the number of trials (lag: 0, 1, 5, or 20), separating the initial presentations (study items) and critical test probes (either an old face or a conjunction lure). A pair of parent faces gave rise to 2 conjunction faces, with the conjunction faces occurring at lags of either 0 and 5 or 1 and 20, where the parent–conjunction lag was determined by the number of trials separating the most recent parent face and the conjunction face (e.g., T. C. Jones & Atchley, 2002; Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996). A pair of parent faces (always constructions) never appeared in adjacent trials; the number of trials separating parent faces (i.e., parent–parent lag) ranged from 1 to 2, with a mean of 1.45 trials. There were 10 items in each condition (old, lag 0, 1, 5, or 20; critical lure, lag 0, 1, 5, or 20) and items occurred in each condition equally often across subjects. (Original faces always served as the test faces.) Also, for conjunction conditions, the order of the parent faces was counterbalanced. Twenty filler faces were used, and these faces were repeated once in the list at a variety of retention intervals. On each trial, a single face was presented for 1,500 msec, followed by a 150-msec blank screen, which was followed by a 2,500-msec response period, during which subjects judged whether the face was "new" or "old" in comparison with any earlier face on a 6-point confidence rating scale (1 = *very sure new*, 6 = *very sure old*). The 2,500-msec response period was held constant so that when a subject responded before the time limit expired, the computer was blank for the remainder of the trial. The ITI was 500 msec.

All of the subjects were specifically instructed, "Your task is to decide how confident you are of whether each face is NEW (not presented earlier in the list) or OLD (presented EXACTLY as it was earlier in the list)." Only the subjects in the exclusion test group were told about the presence of conjunction lures and how an error could be avoided: "Some of the faces will be constructions (i.e., combinations) of other faces. Although these combination faces contain features that you have already seen (in different faces), these combination faces are NEW. If you can remember HOW a face looked different (maybe the hair, ears, chin, and forehead were presented with a different set of eyes, nose, and mouth) on an earlier trial, then you can be certain that the face is NEW." For the unwarned (tacit inclusion) instructions, subjects were told that some test faces would be repeated within the list (i.e., old) but that others would not be repeated (i.e., new).

Results and Discussion

The proportions of "old" responses and "very sure old" responses for old and conjunction conditions are shown in Figures 5A and 5B, respectively. The functions have been corrected for the baseline error rates, which are given in the figure captions. As one would expect, "old" and "very sure old" responses were higher for the *old* conditions than for the conjunction conditions. More importantly, the quadratic trend for the conjunction conditions observed in T. C. Jones and Atchley (2002, 2006)—with a relatively low error rate at the 0-trial lag—was not obtained. (No statistics are needed to verify this observation.) Instead, the "old" response rate, as well as the "very sure old" response rate, was *highest* at the 0-trial lag and decreased as the lag retention interval increased, to produce a forgetting curve. The same pattern was observed for both sets of test instructions, and the conjunction and *old* conditions produced the same pattern. Thus, the data on "old" responses provide no evidence that recollection was used successfully to avoid conjunction errors in the warned group.
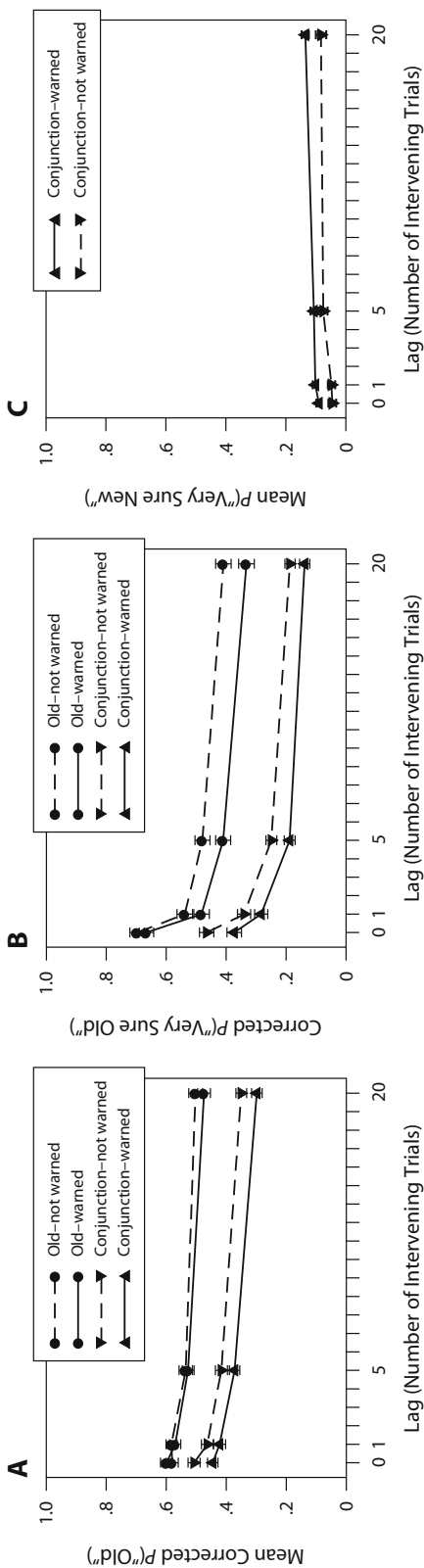
Figure 5. Experiment 2: (A) Corrected proportions of "old" responses for *old* (hits) and conjunctions (false alarms) conditions across lags (the baseline false alarm rates, .33 for the no warning group and .28 for the warning group, have been subtracted). (B) Corrected proportions of confident "old" responses for *old* (hits) and conjunction faces (false alarms) conditions across lags (the baseline false alarm rates, .09 for the no warning group and .07 for the warning group, have been subtracted). (C) Very confident correct rejections of conjunction lures by test group across lags. For all three panels, the error bars represent the standard error for each point. (The standard errors were low and may tend to overlap the symbols indicating the means.)

Figure 5C shows the mean proportions of "very sure new" responses for conjunction faces across the lags for each group. The important finding is the conspicuous absence of a robust proportion of "very sure new" responses for conjunction faces at the zero lag (compare Figures 5C and 1B). The "very sure new" response rate was *lowest* at the zero lag, and the rate increased subtly across lags. Note that there was an effect of lag [$F(3,474) = 8.89$, $MS_e = 0.006$]. This pattern was the same for both groups. The warned group showed a higher rate of "very sure new" responses for conjunction faces than did the unwarned group (.11 vs. .06, respectively) [$F(1,158) = 11.94$, $MS_e = 0.028$], but the warned group also showed a higher rate of "very sure new" responses to entirely new faces than did the unwarned group (.38 vs. .32, respectively) [$F(1,158) = 3.93$, $MS_e = 0.038$]. Thus, the warned group appears simply to have been more conservative with its responses.

Finally, we obtained conjunction–new ROC functions for the unwarned and warned groups for each of the lag retention intervals (see Figures 6A and 6B). Important here is that none of the empirical conjunction–new points or the ROC functions fitted to these points dipped across the unity diagonal, even for the 0-lag condition. Again, the configuration–recollection hypothesis received no support.

To summarize, the "old" response rates, confident correct rejections, and ROC functions converge on the conclusion that subjects were unable to control conjunction errors, even when the features had been presented on the immediately preceding trial (about 3 sec earlier).

EXPERIMENT 3

Demonstrations (e.g., Sinha & Poggio, 1996, 2002) and empirical findings obtained with different types of feature (part old–part new) lures (T. C. Jones et al., 2006) have shown that subjects use the inner and outer features to recognize a face. Nonetheless, one might wonder, as one reviewer did, whether subjects in Experiments 1 and 2 depended primarily on the inner features of a face when making a recognition memory decision. In addition, a context in which old features are juxtaposed with new features (i.e., feature lures) might somehow provide a greater impetus for a successful recall-to-reject strategy. Therefore, in this final experiment, we compared a feature lure condition and a conjunction lure condition over the same short retention intervals as in Experiment 2, and the old set of features (inner or outer) in the feature lure condition was counterbalanced across subjects.

In addition, although subjects in the warned condition of Experiment 2 were told about the conjunctions, they did not see examples of old, conjunction, and new faces. The lack of examples raises the question of how well subjects understood the types of items that they would encounter. Therefore, to be sure that subjects understood the conditions and correct responses, we included a short, open practice test in which a face was shown, then joined with the correct response (with critical lure trials specifically identified), and then followed by the confidence rating scale.
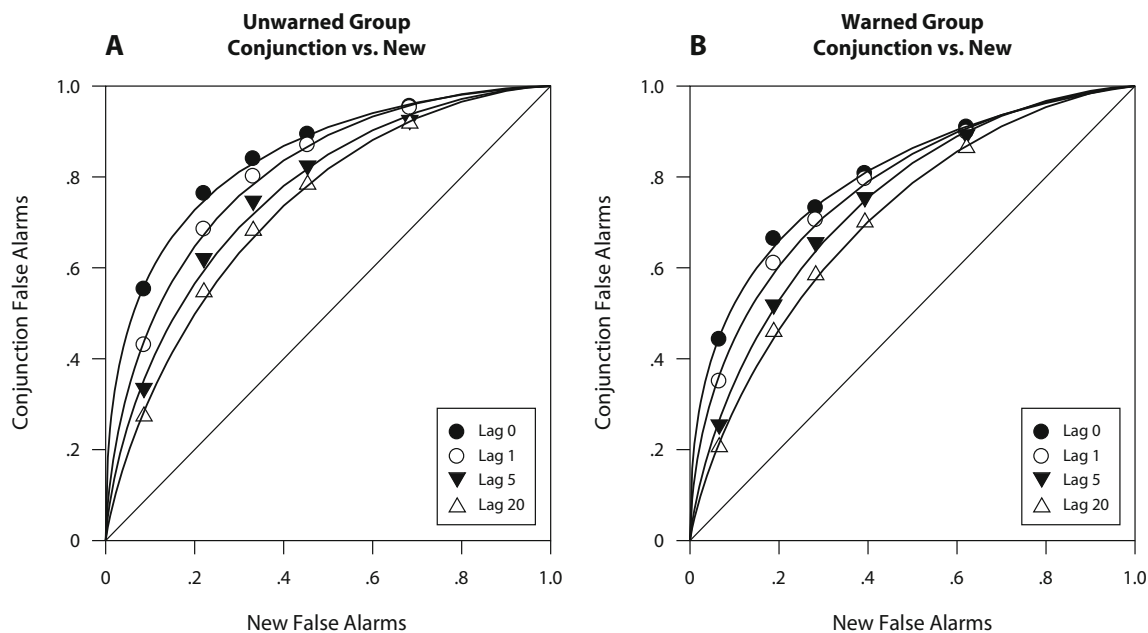
Figure 6. Experiment 2: Conjunction–new ROC functions for the unwarned (A) and warned (B) groups at each lag retention interval.

Method

Subjects. VUW undergraduates ($N = 64$) participated voluntarily or for credit toward a class research requirement. The subjects were assigned randomly to one of two groups ($n = 32$ per group).

Apparatus. The faces were those from Experiment 2, plus 40 additional faces that were used as fillers.

Design and Procedure. There were two groups of subjects. For one group, the critical lures were conjunction lures, but for the other, the critical lures were feature lures. The type of item (old or critical lure) and the retention interval (lag in number of trials) were manipulated within subjects. For the feature lure group, the old half (inner or outer features) of a feature lure was counterbalanced across subjects. For this group, the parent face that occurred first in the list was replaced by a filler face. To avoid potential confusion with regard to the words *old* and *new*, we changed the confidence scale and test instructions so that *same* and *different*, respectively, appeared in their place. An eight-trial, open practice test was included before the actual test. A face was shown for 2,000 msec; then it was joined with the correct response (with critical lure trials specifically identified) for 2,000 msec, which was then followed by the confidence rating scale for up to 2,500 msec. The timing parameters of the actual test were the same as those in Experiment 2.

Results and Discussion

The proportions of "same" and "very sure same" responses for *old* and critical lure conditions are shown in Figures 7A and 7B, respectively. The functions have been corrected for the baseline error rates, which are given in the figure captions. The results reproduced the findings for conjunction lures from Experiment 2 and extended them to feature lure conditions. The hit-rate and error-rate patterns were the same as in Experiment 2, with the highest hit rate and error rate occurring at the 0-trial lag. The corrected feature error rate was lower than the conjunction error rate at all lag retention intervals. A repeated measures ANOVA on the corrected "same" responses was carried out with lag as a within-subjects factor and feature lure type as a between-subjects factor. Although the effect of lag was significant [$F(3,93) = 3.38$, $MS_e = 0.19$], neither the effect of feature lure type [$F(1,31) = 0.27$, $MS_e = 0.01$] nor the lag × feature lure type interaction [$F(1,31) = 1.85$, $MS_e = 0.08$] was significant. A repeated measures ANOVA on the "same" responses across lag conditions with critical lure type (conjunction or feature, i.e., collapsed across feature lure type) as a between-subjects factor gave significant effects of lag [$F(3,186) = 9.60$, $MS_e = 0.25$] and critical lure type [$F(1,62) = 20.16$, $MS_e = 1.79$], but the lag × critical lure type interaction was not significant [$F(3,186) = 0.62$, $MS_e = 0.02$].

Next, we considered the "very sure different" responses (see Figure 7C). As one might expect, the rate of "very sure different" responses was higher for feature lures than for conjunction lures. This finding makes sense in that only one set of features had been seen previously for feature lures, whereas both inner and outer features had been seen for conjunction lures. More important is the pattern of "very sure different" responses across the lag retention interval. For both conjunction and feature lure conditions, the 0-lag condition had the lowest "very sure different" rate, with the confident rejection rate generally increasing as the lag retention interval increased. Therefore, the pattern for conjunction lures reproduced that obtained in Experiment 2; this pattern was also obtained for feature lures.

Finally, we obtained conjunction–new and feature–new ROC functions for each of the lag retention intervals (see Figures 8A and 8B). As in the earlier experiments, none of the empirical (or fitted) ROC functions dipped across

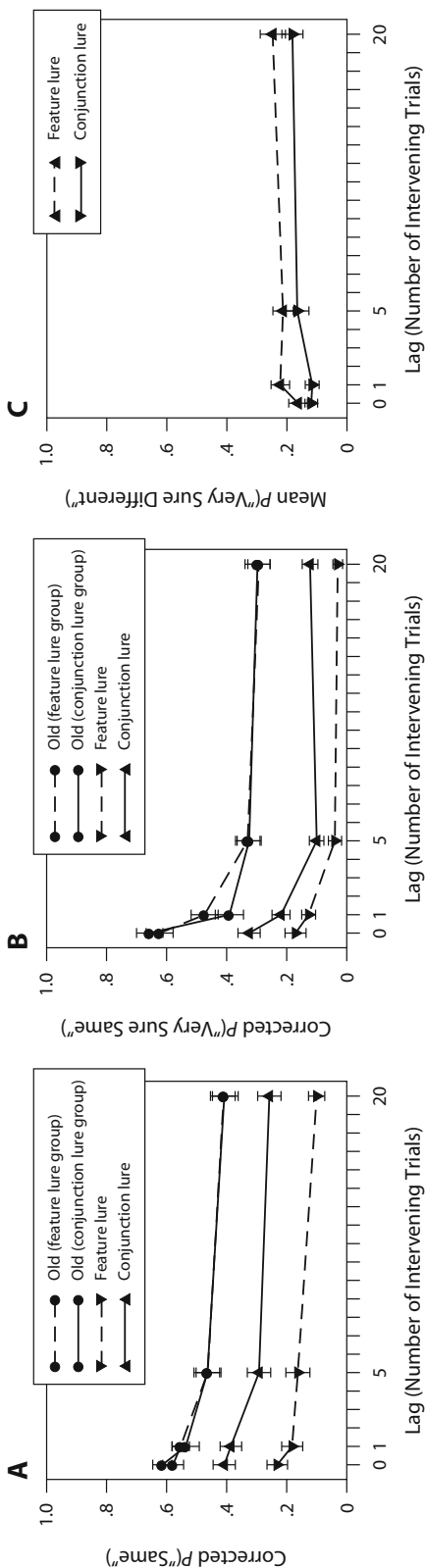**Figure 7. Experiment 3: (A)** Corrected proportions of "same" responses for *old* (hits) and critical lure faces (false alarms) conditions across lags (the baseline false alarm rates, .28 for the feature lure group and .25 for the conjunction lure group, have been subtracted). **(B)** Corrected proportions of confident "same" responses for *old* (hits) and critical lure faces (false alarms) conditions across lags (the baseline false alarm rates, .05 for the feature lure group and .04 for the conjunction lure group, have been subtracted). **(C)** Very confident correction rejections of conjunction lures by test group across lags. For all three panels, the error bars represent the standard error for each point. Many of the old scores, overall (A) and confident (B), overlap for the two groups. The standard errors were low and may tend to overlap the symbols indicating the means.

the unity diagonal, even for the 0-lag condition. None of the measures in this final experiment provided evidence that recollection-based rejections of conjunction or feature lures occurred.

GENERAL DISCUSSION

Subjects in the present experiments had the opportunity to recall the configuration of at least one parent, to avoid misidentifying feature (Experiment 3) or conjunction (Experiments 1–3) faces as "old." Across the three experiments, the pattern of the "old" responses, the proportion of "very sure new" responses, and the ROC functions converged on the conclusion that recollection was not used, or at least not used successfully, to reject conjunction lures of naturalistic faces. The manipulation of test instructions appeared to impact only the overall willingness of subjects to endorse test faces as "old." Study repetition and shorter lag retention intervals led to higher hit rates, but these relatively high hit rates came at a cost: Conjunction error rates were also at their highest after multiple study repetitions and the shortest lag retention. Jacoby (1999) described similar findings with study repetition of words in list discrimination paradigms as "ironic." Consistent with his description, the present results show ironic effects of study repetition and retention interval. Especially ironic is that the proportion of "very sure new" responses was lowest in the conditions (multiple study presentations, very short retention interval) where a recall-to-reject strategy should have been most successful. Finally, none of the conjunction–new or feature–new ROC points or fitted ROC functions dipped below the unity diagonal in a manner indicative of a successful recall-to-reject strategy.

Overall, the present experiments stress the importance of familiarity in face recognition for unknown faces (T. C. Jones et al., 2006; see also Vokey & Read, 1992). This reliance on familiarity is clearly useful but prone to error. This conclusion applies to the situations (multiple study presentations and immediate lag retention interval) in which memory might be thought to be very good, and, in fact, was very good in terms of old–new discrimination. Also, the conjunction errors at the strongest confidence level would appear to be subjectively compelling errors (see also Lampinen et al., 2004; Odegard et al., 2005, who place an emphasis on subjectively compelling errors in their phantom ROC model). Some researchers have concluded that spaced repetition increases recollection for faces (Mäntylä & Cornoldi, 2002) and that configurations support recollection (Mäntylä & Holm, 2005). The present research, however, provides no clear evidence that a recollection process aids in the rejection of a face that shares only some of its parts with a previously studied face. In addition, other researchers have suggested that recollection might somehow exacerbate conjunction errors (Lampinen et al., 2004; Marsh, Hicks, & Davis, 2002), but we have no clear evidence from the present experiments that the recall of facial features exacerbated errors.[5]

Again, all of the research that has promoted the dual-process approach with faces has used the remember–know procedure and has been based on the assumption that re-
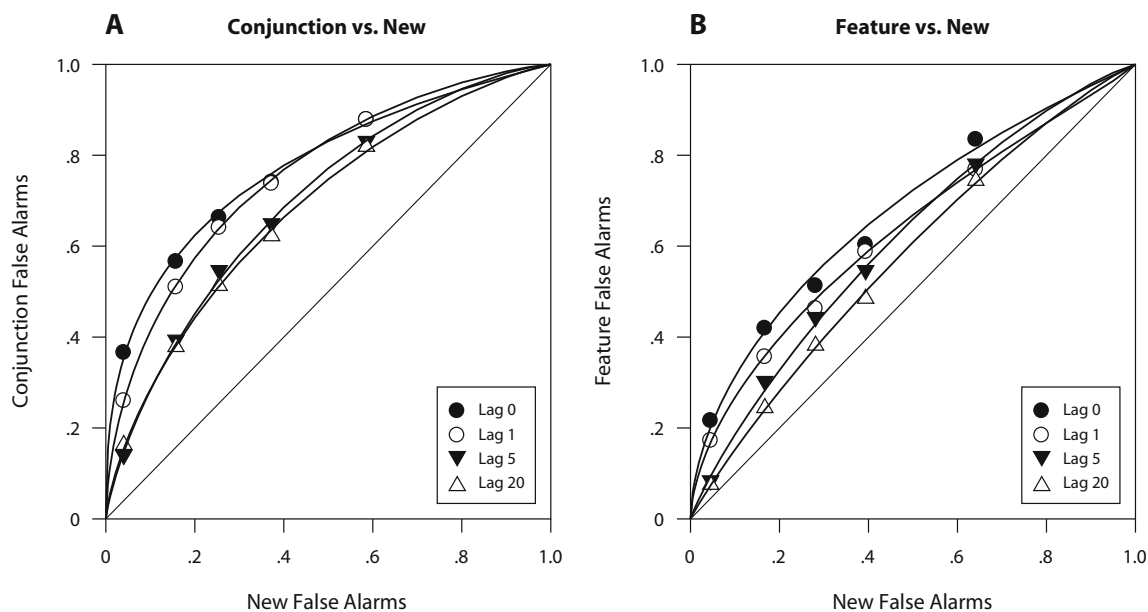
Figure 8. Experiment 2: Critical lure–new ROC functions for the conjunction lure (A) and feature lure (B) groups at each lag retention interval.

member judgments provide a good measure of recollection. Recent research done with other stimuli, however, has indicated that remember judgments may not provide a pure measure of recollection (Dougal & Rotello, 2007; Rotello et al., 2005). Alternatively, we note that the information supporting remember responses in other experiments may not reflect the use of recollection for the face itself. For example, contextual details (e.g., time, place, thoughts, conversations, actions, etc.; cf. Gruppuso, Lindsay, & Masson, 2007; James, 1890) might be recalled instead of the actual global configuration of a face.

A tempting proposal would posit that the difference between old and conjunction faces in the present experiments was based on the successful use of recollection to retrieve configural information in old faces but not in conjunction faces. Such a proposal would generally be consistent with the conclusion of Mäntylä and Holm (2005), who found a significantly higher rate of remember responses for upright human faces than for inverted human faces, upright horse faces, and inverted horse faces. The remember response rates were interpreted as a measure of recollection. In contrast, Mäntylä and Holm found no differences among know judgments and interpreted that result as a measure of familiarity. Thus, they interpreted their findings as showing that configural information supports recollection, but not familiarity, in face recognition memory.

Mäntylä and Holm's (2005) findings for remember and know judgments are certainly of interest, but it is important to note that their interpretation of those findings assumed a relationship of mutual exclusivity for the underlying processes of recollection and familiarity (see also Mäntylä, 1997; Mäntylä & Cornoldi, 2002). Two other

possible relationships between recollection and familiarity consist of redundancy, where familiarity is measured by the hit rate (i.e., the summed rates of remember and know judgments) and independence, where familiarity is measured by the proportion of know judgments divided by one minus the proportion of remember judgments [i.e., familiarity = know/(1 − remember)] (for relevant work, see Arndt & T. C. Jones, 2008; Jacoby et al., 1998; G. V. Jones, 1987; Joordens & Merikle, 1993). Using Mäntylä and Holm's means (estimated from their figures), we estimated familiarity under a redundancy assumption and an independence assumption. In both cases, inversion of faces lowered the measure of familiarity along with the measure of recollection. Thus, either of these alternative relationship assumptions leads to a different conclusion than that drawn by Mäntylä and Holm: Inversion and, by implication, changes in configural information also affect familiarity.

Yonelinas, Kroll, Dobbins, and Soltani (1999) obtained evidence that familiarity contributed to discrimination of old from conjunction face drawings in an upright orientation. Their conclusion was based on a dual-process model (e.g., Yonelinas, 1994, 1997), where recollection is characterized as a high-threshold process, but familiarity is characterized as a continuum process conforming to signal-detection properties. In that model, if familiarity contributes to recognition memory, the ROC function will show some curvilinearity. If recollection is the sole basis for old–conjunction discrimination, the ROC function will be linear. In their study, the old–conjunction ROC function was curvilinear, indicating that familiarity was involved.[6] (Yonelinas et al. [1999] found that recollection was the sole basis of old–conjunction discrimination for

inverted faces, but [1] discrimination was low and [2] this inverted condition is thought to severely disrupt the processing of configural information; see, e.g., Searcy & Bartlett, 1996.)

We do not deny the possibility that recollection contributed to old–conjunction discrimination. However, the present findings can be seen most parsimoniously as consistent with a single-process model, as in global matching models (e.g., Clark & Gronlund, 1996), where memory strength can be construed as familiarity strength, or with a unidimensional model (see Dunn, 2008, for a review). For faces, an episodic face recognition memory model, an autoassociative-general abstract processing system (GAPS) model (Bartlett et al., 2003), is generally consistent with global matching models, because it too can be viewed as being based on familiarity. The autoassociative-GAPS model treats faces holistically, in that it does not extract what are typically viewed as individual features (e.g., nose length, lip thickness). Instead, the model recognizes faces via a pixel-based code. In simulations, the model has overestimated conjunction–new discrimination (i.e., the conjunction effect) and underestimated old–conjunction discrimination, outcomes that, given our present results, may be due to the lack of a configuration parameter. Future development of this model or other global matching models to simulate recognition memory for unknown faces would appear to be worthwhile. One implication of such model development might be that standard featural information may not be required per se.

The present findings stand in stark contrast to those of comparable studies with compound word conjunctions (T. C. Jones, 2005; T. C. Jones & Atchley, 2002, 2006; T. C. Jones & Jacoby, 2001; Lampinen et al., 2004). Of course, the compound words used in earlier studies and the novel faces in the present experiments have obvious differences. One difference is that compound words, as well as their constituents, are known to subjects, whereas our faces are unknown to subjects (although subjects certainly know that a face is a face and a nose is a nose, etc.). Also, we suspect that the constituents for compound word conjunctions may be more salient (i.e., readily processed as units) than the features of faces (see Tanaka & Farah, 1993) and that the role of meaning sometimes may be important for compound words (e.g., Odegard et al., 2005) but not for unknown faces. Finally, it is not clear whether the manipulations that have produced evidence for recollection-based rejections of compound word feature and conjunction lures would extend to other word materials (e.g., where the "feature" constituents are syllables; T. C. Jones, Brown, & Atchley, 2007; Rubin, Van Petten, Glisky, & Newberg, 1999; Underwood & Zimmerman, 1973) or extend to various nonword materials (e.g., T. C. Jones et al., 2007).

Implications for Working Memory/ Short-Term Memory

T. C. Jones and Atchley (2006) noted that, for the 0-trial lag retention interval in their continuous recognition experiment with compound words, a parent word could be retrieved from working memory/short-term memory. In the present Experiments 2 and 3, a parent face had the potential to be retrieved from visual working memory, but the results provide no evidence that working memory was used to control conjunction errors. (Alterations in the procedure might produce different results.) This conclusion agrees with the idea that subjects may not be able to rehearse a complex visual image (Shaffer & Shiffrin, 1972). The outcome is also consistent with recent findings on working memory from a change detection task, where a current issue concerns the role of stimulus complexity and working memory capacity (cf. Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; H. Y. Eng, Chen, & Jiang, 2005). For example, for a memory array comprising upright faces (exclusively), the working memory capacity estimate found by Scolari, Vogel, and Awh (2008) was rather low (1.72 faces). To be clear, Scolari et al. argued that working memory capacity is relatively fixed at about four slots, but that the representations that fill those slots may be fuzzy. From a functional viewpoint, though, it does not matter whether a visual working memory representation was absent or whether it was present but fuzzy in our continuous recognition task: The false recognition errors appear to have been out of reach of a controlled retrieval process.

Conclusion

None of the present experiments with unknown faces produced evidence that face conjunction errors could be controlled by using recollection to retrieve a studied configuration, despite the inclusion of conditions designed to allow for such control. Overall, the results emphasize the strong role of familiarity in episodic face recognition, including a role in old–conjunction discrimination of upright faces. The results are consistent with a single-process model of recognition memory, but further research should probe whether a dual-process approach might sometimes be required to account for face recognition memory results. The dual-process approach might be helpful in accounting for context effects (e.g., Gruppuso et al., 2007), but the present results bring into question whether it is necessary in accounting for recognizing a face itself.

REFERENCES

ALVAREZ, G. A., & CAVANAGH, P. (2004). The capacity of visual short-term memory is set both by information load and by number of objects. *Psychological Science*, **15**, 106-111.

ARNDT, J., & JONES, T. C. (2008). Elaborative processing and conjunction errors in recognition memory. *Memory & Cognition*, **36**, 899-912.

AWH, E., BARTON, B., & VOGEL, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, **18**, 622-628.

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

BARTLETT, J. C., SEARCY, J. H., & ABDI, H. (2003). What are the routes to face recognition? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes. Advances in visual cognition* (pp. 21-47). Oxford: Oxford University Press.

BASTIN, C., & VAN DER LINDEN, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology*, **17**, 14-24.

CABEZA, R., & KATO, T. (2000). Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, **11**, 429-433.

CAREY, S., & DIAMOND, R. (1977). From piecemeal to configurational representation of faces. *Science*, **195**, 312-314.

CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.

DIANA, R. A., REDER, L. M., ARNDT, J., & PARK, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, **13**, 1-21.

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

DOUGAL, S., & ROTELLO, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review*, **14**, 423-429.

DUNN, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review*, **111**, 524-542.

DUNN, J. C. (2008). The dimensionality of the remember–know task: A state-trace analysis. *Psychological Review*, **115**, 426-446.

ENG, H. Y., CHEN, D., & JIANG, Y. (2005). Visual working memory for simple and complex stimuli. *Psychonomic Bulletin & Review*, **12**, 1127-1133.

ENG, J. (n.d.). ROC analysis: Web-based calculator for ROC curves. Retrieved December 7, 2007, from www.jrocfit.org.

FARAH, M. J., WILSON, K. D., DRAIN, M., & TANAKA, J. N. (1998). What is "special" about face perception? *Psychological Review*, **105**, 482-498.

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

GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics.* Huntington, NY: Krieger.

GRUPPUSO, V., LINDSAY, D. S., & MASSON, M. E. J. (2007). I'd know that face anywhere! *Psychonomic Bulletin & Review*, **14**, 1085-1089.

THE HAWKEYE (1963). Iowa City: State University of Iowa.

HEATHCOTE, A., RAYMOND, F., & DUNN, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory & Language*, **55**, 495-514.

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

JACOBY, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 3-22.

JACOBY, L. L., JONES, T. C., & DOLAN, P. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, **5**, 705-709.

JAMES, W. (1890). *Principles of psychology: Vol. 1.* London: Macmillan.

JONES, G. V. (1987). Independence and exclusivity among psychological processes: Implications for the structure of recall. *Psychological Review*, **94**, 229-235.

JONES, T. C. (2005). Study repetition and the rejection of conjunction lures. *Memory*, **13**, 499-515.

JONES, T. C., & ATCHLEY, P. (2002). Conjunction error rates on a continuous recognition memory test: Little evidence for recollection. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 374-379.

JONES, T. C., & ATCHLEY, P. (2006). Conjunction errors, recollection-based rejections, and forgetting in a continuous recognition task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **32**, 70-78.

JONES, T. C., BARTLETT, J. C., & WADE, K. A. (2006). Nonverbal conjunction errors in recognition memory: Support for familiarity but not for feature bundling. *Journal of Memory & Language*, **55**, 138-155.

JONES, T. C., BROWN, A. S., & ATCHLEY, P. (2007). Feature and conjunction effects in recognition memory: Toward specifying familiarity for compound words. *Memory & Cognition*, **35**, 984-998.

JONES, T. C., & JACOBY, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory & Language*, **45**, 82-102.

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

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

KROLL, N. E. A., KNIGHT, R. T., METCALFE, J., WOLF, E. S., & TULVING, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory & Language*, **35**, 176-196.

LAMPINEN, J. M., ODEGARD, T. N., & NEUSCHATZ, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 332-342.

LIGHT, L. L., PATTERSON, M. M., CHUNG, C., & HEALY, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition*, **32**, 1182-1193.

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

MÄNTYLÄ, T. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1203-1216.

MÄNTYLÄ, T., & CORNOLDI, C. (2002). Remembering changes: Repetition effects in face recollection. *Acta Psychologica*, **109**, 95-105.

MÄNTYLÄ, T., & HOLM, L. (2005). Remembering parts and wholes: Configural processing in face recollection. *European Journal of Cognitive Psychology*, **17**, 753-769.

MARSH, R. L., HICKS, J. L., & DAVIS, T. T. (2002). Source monitoring does not alleviate (and may exacerbate) the occurrence of memory conjunction errors. *Journal of Memory & Language*, **47**, 315-326.

MCELREE, B., DOLAN, P. O., & JACOBY, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 563-582.

MCKONE, E., & PEH, Y. X. (2006). Memory conjunction errors for realistic faces are consistent with configural processing. *Psychonomic Bulletin & Review*, **13**, 106-111.

ODEGARD, T. N., LAMPINEN, J. M., & TOGLIA, M. P. (2005). Meaning's moderating effect on recollection rejection. *Journal of Memory & Language*, **53**, 416-429.

RAKOVER, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology*, **93**, 1-30.

RAKOVER, S. S., & TEUCHER, B. (1997). Facial inversion effects: Parts and whole relationship. *Perception & Psychophysics*, **59**, 752-761.

REINITZ, M. T., LAMMERS, W. J., & COCHRAN, B. P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, **20**, 1-11.

REINITZ, M. T., MORRISSEY, J., & DEMB, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 161-168.

ROTELLO, C. M., MACMILLAN, N. A., REEDER, J. A., & WONG, M. (2005). The *remember* response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*, **12**, 865-873.

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

RUBIN, S. R., VAN PETTEN, C., GLISKY, E. L., & NEWBERG, W. M. (1999). Memory conjunction errors in younger and older adults: Event-related

potential and neuropsychological data. *Cognitive Neuropsychology*, **16**, 459-488.

SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002a). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.

SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002b). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

SCOLARI, M., VOGEL, E. K., & AWH, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychonomic Bulletin & Review*, **15**, 215-222.

SEARCY, J. H., & BARTLETT, J. C. (1996). Inversion and processing of component and spatial–relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 904-915.

SEARCY, J. H., BARTLETT, J. C., & MEMON, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition*, **27**, 538-552.

SHAFFER, W., & SHIFFRIN, R. M. (1972). Rehearsal and storage of visual information. *Journal of Experimental Psychology*, **92**, 292-296.

SHEPARD, R. N., & TEGHTSOONIAN, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, **62**, 302-309.

SINHA, P., & POGGIO, T. (1996). I think I know that face . . . *Nature*, **384**, 404.

SINHA P., & POGGIO, T. (2002). "United" we stand. *Perception*, **31**, 133.

SQUIRE, L. R., WIXTED, J. T., & CLARK, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, **8**, 872-883.

TANAKA, J. W., & FARAH, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, **46A**, 225-245.

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

UNDERWOOD, B. J., & ZIMMERMAN, J. (1973). The syllable as a source of error in multisyllable word recognition. *Journal of Verbal Learning & Verbal Behavior*, **12**, 701-706.

VOKEY, J. R., & READ, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, **20**, 291-302.

WAIS, P. E., WIXTED, J. T., HOPKINS, R. O., & SQUIRE, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron*, **49**, 459-466.

XU, J., & MALMBERG, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition*, **35**, 526-544.

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

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

YONELINAS, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1415-1434.

YONELINAS, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517.

YONELINAS, A. P., KROLL, N. E. A., DOBBINS, I. G., LAZZARA, M., & KNIGHT, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember–know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, **12**, 323-339.

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

NOTES

1. For the sake of clarity, in the present research, we examined recognition of the whole head (e.g., Bartlett, Searcy, & Abdi, 2003; T. C. Jones et al., 2006; Reinitz et al., 1992; Sinha & Poggio, 1996, 2002) instead of just the inner features (e.g., Cabeza & Kato, 2000; McKone & Peh, 2006). Also, for our stimuli, most of the features (eyes, nose, mouth, jaw, ears, hair) were exact copies from study faces to conjunction faces. Because the size of the eyes, nose, and mouth varied from individual to individual in the photographs, the size of the forehead, cheeks, and chin was altered subtly when the inner features were pasted over the original features.

2. In this article, we attempt to be clear with our usage of *recall*, *recollection*, *recall-to-reject*, and *recollection-based rejection*. Our intention is to use *recall* as a more empirical outcome, *recollection* as a theoretical process, *recall-to-reject* as a strategy, and *recollection-based rejection* to describe how the process recollection has been used (i.e., as the basis for rejecting a critical lure).

3. Typically, confidence ratings are made for both *old* and *new* responses, with hits plotted against false alarms across cumulative levels of confidence. The leftmost point represents the most conservative (highest) level of confidence for "old," with each subsequent point (to the right) including the data for the next lowest level of confidence. Excluding the (cumulative) point of lowest confidence (where hits = 1.0 and false alarms = 1.0), the number of empirical points is 1 minus the number of confidence levels. Chance performance (e.g., hits = false alarms) is represented by a diagonal line from the origin to the 1.0–1.0 corner. ROC functions for discrimination of old from new items may be empirical or may be fitted to the empirical points; these functions are typically curvilinear and are bowed up and to the left. The ROC function may be asymmetrical or symmetrical relative to the unity diagonal, although it is usually asymmetrical for normal subjects (cf. Wais, Wixted, Hopkins, & Squire, 2006; Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998). Different models can be used to produce an ROC function, and the ROC function can vary depending on a given model's parameters. All of the ROC functions presented later in this article were generated by the standard, equal variance signal-detection model.

For a psychophysical approach to ROC functions, one would obtain many observations to achieve very stable ROC functions at the subject level (cf. work done in psychophysics or work done in the animal learning literature, or see Heathcote, Raymond, & Dunn, 2006, for an example in recognition memory). ROC functions from individuals, as well as the averaged ROC functions, can be compared with predictions from different models, and statistics can be done to test different models or model parameters. Our present work was not motivated to test particular parameters of any quantitative model or to provide model-fitting comparisons of any formal, quantitative models.

4. The ROC functions from the pooled frequencies in all of the present conjunction–new conditions provide for an easy, qualitative, visual analysis. Out of curiosity, however, we did obtain an average conjunction–new ROC function based on the ROC functions of 72 out of 80 subjects in one of the conditions in Experiment 2 and compared it with the ROC function from the pooled frequencies of those 72 subjects. The 95% confidence intervals of the two functions overlapped at all points, with the pooled ROC function below that of the average ROC from individuals. Not counting the endpoints (0.0, 0.0; 1.0, 1.0), 26 points were used for the ROC functions; 10 of the ROC points from the pooled data fell outside the lower bound of the 95% confidence interval for the ROC function averaged from individuals, with a mean difference from the lower bound of only .009. The slope of the standardized ROC function (*z*-ROC) averaged from the individual subjects (1.11) was actually higher than that from the pooled frequencies (0.95), but there were extreme scores that clearly contributed to the difference. The bottom line is that these two ROC functions were quite similar, particularly for a qualitative, visual analysis.

5. The only evidence toward this idea comes from the shape of the ROC functions. The slope of an ROC function for standardized scores (*z*-ROC) provides a measure of symmetry of the ROC curve. A perfectly symmetrical ROC curve will have a *z*-ROC slope of 1.0. In the standard signal-detection model, an asymmetry (e.g., slope markedly lower than 1.0) may be interpreted simply to mean that memory is strong (e.g., Squire, Wixted, & Clark, 2007; Wais et al., 2006). However, in Yonelinas's (1994, 1997, 1999) approach, a reliance solely on familiarity will produce a symmetrical ROC function, a reliance solely on recollection will produce a linear ROC function, and the use of both familiarity and recollection will produce a curvilinear ROC function that is asymmetrical to some degree. Thus, an asymmetry can be taken to mean that recollection supports recognition, in part.

Again, using the group frequencies, we obtained the *z*-ROC slopes for conjunction–new discrimination in all (12) conjunction–new conditions

across the first two experiments (where there were many observations). The inclusion/unwarned conditions should bias our conclusion toward finding evidence of the use of recollection because, under these conditions, recollection could be used to accept a conjunction face as "old"; thus, we were liberal with our approach to identifying the use of recollection in conjunction conditions). The mean slope was 0.95, which can be interpreted as quite symmetrical. Thus, the conjunction errors appear either to be relatively weak in terms of memory strength or to be based on familiarity (or both). The relatively weak conditions (one study presentation in Experiment 1, and 1-, 5-, and 20-trial retention intervals in Experiment 2) produced symmetrical conjunction–new ROC functions, with a mean z-ROC slope of 1.01, whereas the relatively strong memory conditions (i.e., multiple study presentations in Experiment 1, 0-trial retention interval in Experiment 2) produced mildly asymmetrical conjunction–new ROC functions, with z-ROC slopes of 0.83 and 0.84, respectively, averaged across inclusion/exclusion and warned/unwarned groups. The approach taken by Yonelinas (1994, 1997, 1999) and Lampinen et al. (2004) would encourage the interpretation that recollection may have exacerbated errors in the strong conditions but not the others, but, as far as we can tell, there is nothing inherent in these models that would predict this outcome. (Such an a priori prediction of this pattern would seem somewhat unlikely.)

6. We used Yonelinas's (1994, 1997) model expressed in Excel and obtained curvilinear ROC functions for old–conjunction and old–feature discrimination. There are parameters for recollection and familiarity, and the Solver routine of Excel produces a y-intercept (a measure of recollection), a fixed d' value, and multiple values for the decision criterion (C). These values and the fitted ROC are determined by minimizing the summed squared differences of predicted versus observed hits and predicted versus observed false alarms. Overall, Yonelinas et al.'s (1999) finding for upright face drawings extended to our naturalistic faces. All but three of the old–conjunction ROC functions were curvilinear, and the z-ROC slope for several of these functions indicated that discrimination was based heavily on familiarity according to Yonelinas's (1994, 1997) rationale. The three old–conjunction ROC functions that were not curvilinear (i.e., they were linear) came from the three zero-trial lag conjunction conditions of Experiments 2 and 3. We are cautious about the interpretation of these linear ROC functions, because the empirical points were bunched in the upper right half of the plot, making a linear ROC function much more likely to be obtained with that model.