Age differences in accuracy and choosing in eyewitness identification and face recognition

JEAN H. SEARCY and JAMES C. BARTLETT
University of Texas at Dallas, Richardson, Texas

and

AMINA MEMON
University of Southampton, Southampton, England

Studies of aging and face recognition show age-related increases in false recognitions of new faces. To explore implications of this false alarm effect, we had young and senior adults perform (1) three eyewitness identification tasks, using both target present and target absent lineups, and (2) an old/new recognition task in which a study list of faces was followed by a test including old and new faces, along with conjunctions of old faces. Compared with the young, seniors had lower accuracy and higher choosing rates on the lineups, and they also falsely recognized more new faces on the recognition test. However, after screening for perceptual processing deficits, there was no age difference in false recognition of conjunctions, or in discriminating old faces from conjunctions. We conclude that the false alarm effect generalizes to lineup identification, but does not extend to conjunction faces. The findings are consistent with age-related deficits in recollection of context and relative age invariance in perceptual integrative processes underlying the experience of familiarity.

A well-known generalization about aging and memory is that tests of recall show strong age-related deficits, whereas tests of recognition often do not (see, e.g., Craik & Jennings, 1992; Kausler, 1994). However, this generalization admits of exceptions, and the present research is focused on one: recognition memory for unfamiliar faces. Older adults perform relatively more poorly in distinguishing strangers' faces viewed previously in some context from entirely new faces not ever viewed before. The effect is clearly relevant to contemporary theories of memory (and false memory) in old age (Schacter, Koutstaal, & Norman, 1997), and it carries potential applied implications regarding the treatment of older eyewitnesses to crime (Bornstein, 1995). Hence, age-related deficits in face recognition memory are important to examine.

The existence of these deficits should not be in doubt. They have been found with a variety of methods, including free choice, forced choice, and delayed matching-to-sample testing (e.g., Crook & Larrabee, 1992; Grady et al., 1995; Smith & Winograd, 1978), with both study–test and continuous recognition designs (Ferris, Crook, Clark, McCarthy, & Raye, 1980), and in both incidental and intentional learning tasks (Smith & Winograd, 1978). A striking feature of such deficits is a characteristic pattern

of hits and false alarms observable in free choice tests. Hit rates are generally similar for young and older subjects, averaging .82 versus .81, respectively, across the 12 published experiments reporting this measure that are listed in Table 1. In contrast, false alarm rates in young adult groups are consistently lower than those in older groups, averaging .20 versus .40, respectively, across the 12 experiments.[1]

Two recent findings offer clues about the nature of the false alarm effect. First, in Bartlett and Fulton (1991), older but not younger adults showed positive correlations between subjective ratings of the familiarity of unknown faces and hit and false alarm rates for the same faces in old/new recognition tests. Second, in Bartlett, Strater, and Fulton (1991), older but not younger adults showed inflated rates of false *famous* judgments to nonfamous faces that they had seen 1 week prior to the test. Similarly, older but not younger adults showed high rates of false *recent* judgments to *nonrecent* faces that they had seen 1 week before the test. These and related findings (e.g., those of Dywan & Jacoby, 1990) suggested a *context recollection* hypothesis that can be framed in terms of Mandler's (1980) distinction between perceived familiarity and retrieval (i.e., recollection).

Although Mandler (1980) developed his theory in the domain of verbal memory, he illustrated it by citing the experience of recognizing that a face is familiar without being able to identify it. According to Mandler, this experience occurs because familiarity is based solely on the perceived integration or organization of the intra-item sensory and perceptual elements constituting the stimulus, apart from the context(s) in which it has been en-

Table 1
Hit and False Alarm Rates by Young Adults and Seniors in Face Recognition

Article	Method	Hit Rate		False Alarm Rate	
		Young	Senior	Young	Senior
Smith & Winograd (1978)	Standard, friendliness- or nose-rating task at study	.80	.79	.12	.30
Mason (1986)	Study list of young and elderly faces and names followed by recognition and forced-choice naming	.89	.88	.14	.67
Bartlett & Leslie (1986)	Most "old" faces changed from study in pose				
	Experiment 1: Study faces shown in 1 or 4 poses	.81	.83	.26	.38
	Experiment 2: Study faces shown in 4 poses	.82	.90	.16	.34
Bartlett, Leslie, Tubbs, & Fulton (1989)	Most "old" faces changed from study				
	Experiment 1: Old/New Judgments	.87	.88	.21	.43
	Experiment 2: Old/Type-of-Change Judgments	.90	.87	.29	.58
Bartlett & Fulton (1991), Experiment 2	"Old" faces were identical to study items or changed in expression	.82	.80	.28	.40
Fulton & Bartlett (1991)	"Old" faces were identical to study items or changed in expression	.84	.84	.29	.46
Bäckman (1991), Experiment 2 69–76 yr. senior groups	Study list and test had famous faces (only data from nonfamous faces shown)	.80	.67	.07	.17
Flicker, Ferris, Crook, & Bartus (1990)	Young adult faces. Continuous recognition, with items repeated after 0- to 300-sec lag	.77	.78	.14	.24
Ferris, Crook, Clark, McCarthy, & Raye (1980)	Young adult faces. Continuous recognition task with lag of .5 to 4 min plus 40-min delayed test	.67	.65	.25	.43
Crook & Larrabee (1992)	As in Ferris et al. (1980), but with variously aged faces	.79	.80	.21	.34
M		.82	.81	.20	.40

Note—Unless otherwise noted, faces were those of nonfamous males and females of various ages.

countered. In contrast, identification depends on recollection of context, which itself depends on elaborative information concerning relations among stimuli and between stimuli and context (see also Read, 1995).

Extending Mandler's (1980) theory, we developed the hypothesis that older persons suffer deficits in recollection of contextual information (cf. Craik & Jennings, 1992; Light, 1991), but *not* in the perceptual integrative processes underlying the detection of familiarity of a previously viewed face (cf. Jacoby, Jennings, & Hay, 1996; Rabinowitz, 1984). This pattern of a deficit in recollecting context and age-invariant familiarity processes leads older persons to rely more than younger persons on familiarity as a basis for recognition decisions (Bartlett, 1993). The result is high rates of false alarm errors with lures that are high in perceived familiarity. Since faces form a highly homogenous set of stimuli with a generally high level of interitem similarity (Diamond & Carey, 1986), even entirely new faces in a recognition test often evoke a strong sense of perceived familiarity because they resemble faces seen in life (e.g., on the street that day). Hence, older persons are highly susceptible to false alarm errors in response to new faces.

Additional support for the context recollection hypothesis derives from several studies of old/new recognition with nonfacial materials. As is shown in Table 2, each of these studies used *familiarized* lures that were presented to subjects outside the study context prior to appearing in the recognition test. Such familiarized lures evoked more false alarms among older subjects ($M = .30$) than among young adults ($M = .17$) across seven relevant experiments. Note that control or *nonfamiliarized* lures

evoked fewer false alarms and showed, on the average, smaller age differences.

Notwithstanding this support, data also suggest two viable alternatives to the context recollection hypothesis. One such alternative is suggested by studies using recognition tests in which some of the lures are contrived or selected to resemble study items perceptually and/or conceptually. As is shown in Table 3, false alarm rates for such resembling lures averaged .35 for elderly subjects versus .20 for young subjects across 21 experiments using various stimuli. These findings suggest an age-related deficit in making fine mnemonic discriminations, perhaps deriving from incomplete or partial learning of stimuli (Smith, 1975), and/or an encoding or retrieval bias favoring "gist" over specific detail information (Bartlett & Leslie, 1986; Koutstaal & Schacter, 1997). Such a deficit would likely cause problems with faces, a class of stimuli whose interitem similarity is known to be high.

Another alternative is suggested by studies showing age-related increases in false alarm rates in tests requiring memory for novel visuospatial information outside the domain of faces. Two of these studies (Park, Puglisi, & Sovacool, 1984; Smith, Park, Cherry, & Berkovsky, 1990) obtained a reliable, albeit small, age difference in false alarms but not hits in recognition memory for more and less complicated scenes. Another relevant study by Schacter, Cooper, and Valdiserri (1992) used possible (spatially coherent) and impossible (spatially incoherent) drawings of novel objects and also used a familiarization manipulation (see Table 2). In this study, the age-related increase in false alarm rates was as strong with nonfamiliarized lures as it was with familiarized lures.[2]

Table 2
Hit and False Alarm Rates by Young Adults and Seniors in Recognition of Familiarized and Nonfamiliarized Test Items

Article	Study Materials	Familiarization Procedure	Hit Rates Familiarized Young	Hit Rates Familiarized Senior	Hit Rates Nonfamiliarized Young	Hit Rates Nonfamiliarized Senior	FA Familiarized Young	FA Familiarized Senior	FA Nonfamiliarized Young	FA Nonfamiliarized Senior
Koriat, Ben-Zur, & Sheffer (1988) Experiment 3	Words	Every old and new item in the recognition test was repeated after 7–31 words	.81	.70	.84	.70	.17	.26	.11	.13
Jennings & Jacoby (1997)	Words	Experiment 1A: Lures repeated at lags of 4–48 intervening items	—	—	.63	.54	.08	.25	.08	.09
		Experiment 2: Lures repeated at lags of 0–12 intervening items; exclusion condition	—	—	.63	.66	.07	.20	.10	.15
Schacter, Koutstaal, Johnson, Gross, & Angell (1997)	Objects seen in video followed by test using verbal-description cues	Experiment 1: Viewing object photos one or three times 20 min after video; test was 2 days later	.96	.91	.79	.65	.13	.21	.07	.10
		Experiment 2: Viewing of object photos 2 weeks after video; test was 20 min later	.88	.69	.81	.49	.22	.40	.22	.28
Park, Puglisi, & Smith (1986), Experiment 1	Normal and high-contrast photos, drawings	The initial recognition test was repeated 4 weeks later	.60	.55	.77	.78	.23	.36	.08	.11
Schacter, Cooper, & Valdiserri (1992), Experiment 1	Possible and impossible novel-object drawings	Recognition preceded by a priming test including old and new drawings	.62	.53	.69	.65	.29	.41	.27	.45
M			—	—	.74	.64	.17	.30	.13	.19

Table 3
Hit and False Alarm Rates by Young Adults and Seniors in Studies of Recognition With Resembling and Control Lures

Article	Study Material	Type of Resembling Lure	Hit Rates Young	Hit Rates Senior	FA Resembling Young	FA Resembling Senior	FA Control Young	FA Control Senior
Smith (1975)	Words	Synonyms, orthographically resembling	.91	.89	.06	.09	.03	.03
Rankin & Kausler (1979)	Words	Synonyms, rhymes	.96	.77	.08	.14	.04	.04
Norman & Schacter (1997)	Words	Associates of sets of study words (Experiment 1)	.79	.75	.71	.85	.12	.26
Reder, Wible, & Martin (1986)	Short stories	More and less plausible inferences	.84	.88	.29	.54	—	—
Hashtroudi, Johnson, Vnek, & Ferguson (1994)	Lines from a play	Paraphrases of lines from the script	.84	.84	.10	.19	—	—
Till, Bartlett, & Doyle (1982)	Scenic photographs	Different parts of same scenes	.69	.64	.17	.19	.04	.06
Bartlett, Till, Gernsbacher, & Gorman (1983), Experiment 1	Scenic photographs	(1) Left/right reversals	.91	.79	.34	.47	.05	.11
		(2) Verbally similar photos	.86	.78	.03	.08	.01	.04
Pezdek (1987)	Complex line drawings	(1) Elaborated versions of drawings	.86	.77	.10	.26	—	—
		(2) Simplified versions of drawings	.84	.73	.27	.55	—	—
Trahan, Larrabee, & Levin (1986)*	Objects of various classes	New items from the studied classes	.96	.92	.20	.35	.03	.05
Trahan, Larrabee, & Quintana (1990)*	Designs of various classes	New items from the studied classes	.90	.90	.16	.25	—	—
Koutstaal & Schacter (1997)†	Colored pictures of objects and animals from small to large classes	Experiment 1: New items from studied classes	.78	.76	.27	.51	.03	.11
		Experiment 2: New items from studied classes	.82	.77	.25	.54	.05	.10
		Experiment 3: New items from studied classes	.84	.77	.21	.48	.05	.22
Cohen & Faulkner (1989)	Actions with objects	Recombinations of actions and objects	.82	.82	.15	.36	—	—
Kroll, Knight, Metcalfe, Wolf, & Tulving (1996)	(1) 2-syllable words	Recombinations of parts of study items	.88	.79	.10	.14	.05	.10
	(2) Schematic face drawings	Recombinations of parts of study items	.92	.81	.40	.67	.03	.06
	(3) Cartoon faces	Recombinations of parts of study items	.96	.96	.07	.31	.03	.00
Chalfonte & Johnson (1996), Experiment 3B	Colored object drawings	Recombinations of drawings and colors						
	(1) Incidental learning		.54	.56	.18	.25	—	—
	(2) Intentional learning		.77	.53	.05	.10	—	—
M			.84	.78	.20	.35	—	—

*Proportions are estimates. †Proportions averaged across small, medium, and large classes.

We conclude that age differences in falsely recognizing faces might reflect deficits in (1) making fine discriminations in memory, (2) processing novel visuospatial information, and/or (3) recollecting context (the context recollection hypothesis). These three accounts are not mutually exclusive, and our main objective in Experiment 1 was to test an interesting and practically important implication that is shared by all three. Experiment 2 addressed a counterintuitive prediction that is unique to the context recollection view, distinguishing it from the others.

EXPERIMENT 1
Age-Related Deficits in Lineup Identification

All three hypotheses under consideration are meant to apply beyond the domain of standard laboratory tests of face recognition, extending to a broad range of tasks in which the processes that they highlight are plausibly involved. One such task is the lineup task used in eyewitness identification research. Many studies have simulated aspects of the lineup situation by presenting subjects with a staged event (live or via video), followed by a lineup task requiring a judgment of whether any of several alternative faces—all matching the same general description—appeared, or played some specified role, in the previous event. The rate of false identifications is generally high, an effect that is known to extend beyond the laboratory: False identification has been found in several studies to be the largest source of unjust convictions (Wells & Seelau, 1995). Clearly, it is both practically and theoretically important to determine whether older eyewitnesses are even more prone to false identifications than are young adults in the lineup task.

An age-related increase in false identifications is predicted by the context recollection hypothesis, since recollection of context is known to be important for lineup identification (Read, 1994). It also is predicted by the hypotheses of age differences in making fine discriminations (lineup foils are chosen to be similar to the target), as well as the hypothesis of a visuospatial memory deficit (faces clearly qualify as visuospatial stimuli). It was not, however, a foregone conclusion, because the few prior studies addressed to the issue have been inconsistent in their findings (Adams-Price, 1992; O'Rourke, Penrod, Cutler, & Stuve, 1989; Scogin, Calhoon, & D'Errico, 1994; Yarmey, 1996; Yarmey & Kent, 1980; Yarmey & Rashid, 1981, cited in Yarmey, 1984). Our aim was to obtain more definitive data on this important issue.

An additional issue that we addressed in this experiment concerns the *bystander effect*, which occurs when a witness mistakenly identifies a lineup member who originally was present at the crime scene, but who was not the actual culprit. Read (1994) argued that the effect might be explained as a negative effect of perceived familiarity on recognition judgments, also pointing out that the effect is ephemeral in studies with young adults.

If Read's (1994) analysis is accurate, the context recollection hypothesis makes an interesting prediction: The bystander effect should be more pronounced among older persons than among young adults. Since neither alternative hypothesis under consideration can make this prediction, we designed one of our three lineup tasks specifically to test it.

A third issue that we examined concerned the possibility that age-related deficits in perceptual processing of faces might affect performance in the lineup task. A well-known test of perceptual processing deficits is the Benton Face Recognition Test (BFRT; see Benton, Sivan, Hamsher, Varney, & Spreen, 1994), a face-matching test that does not involve a memory load. Samples of normal seniors perform below young adults on the BFRT, and a substantial minority of persons 70 years and older show significant impairment (Benton, Eslinger, & Damasio, 1981). In that the BFRT has been found to be correlated with lineup identification and face recognition among healthy young adults (e.g., Hosch, Bothwell, Sporer, & Saucedo, 1990; and see Hosch, 1994), it is essential to determine whether age-related differences in these tasks reflect an increased incidence in old age of the perceptual deficits measured by the BFRT. Such a moderating effect would not be inconsistent with the context recollection hypothesis or others we have considered. However, any such hypothesis that attributes age differences in false alarm errors to processes beyond those of perception must make the following prediction: There should be age-related increases in false identifications, even when persons with low BFRT scores are removed from the elderly sample.

We also addressed metamemory. Self-appraisal of memory has been studied by researchers both in cognitive aging and in eyewitness identification, although in somewhat different ways. The focus among cognitive aging researchers has been on self-appraisals of everyday performance and knowledge of memory tasks and processes. The dominant concern in the eyewitness field has been with confidence in identification (and other) decisions, usually measured retrospectively. We decided to combine these two approaches, administering a self-report instrument based on the Memory Self-Efficacy Questionnaire (Berry, West, & Dennehey, 1989), but tailored to probe prospective confidence along with retrospective confidence in lineup decisions.

Method

Figure 1 summarizes the design and procedure, which are detailed in the following sections.

Subjects

The senior group (ages 60–80, $M = 70$) consisted of 75 individuals, 56% female, recruited through newspaper ads, group meetings, and public notices, who were paid for participation. The young group (ages 18–30, $M = 24$), consisted of 77 University of Texas at Dallas undergraduates, 56% female, who received credit toward the
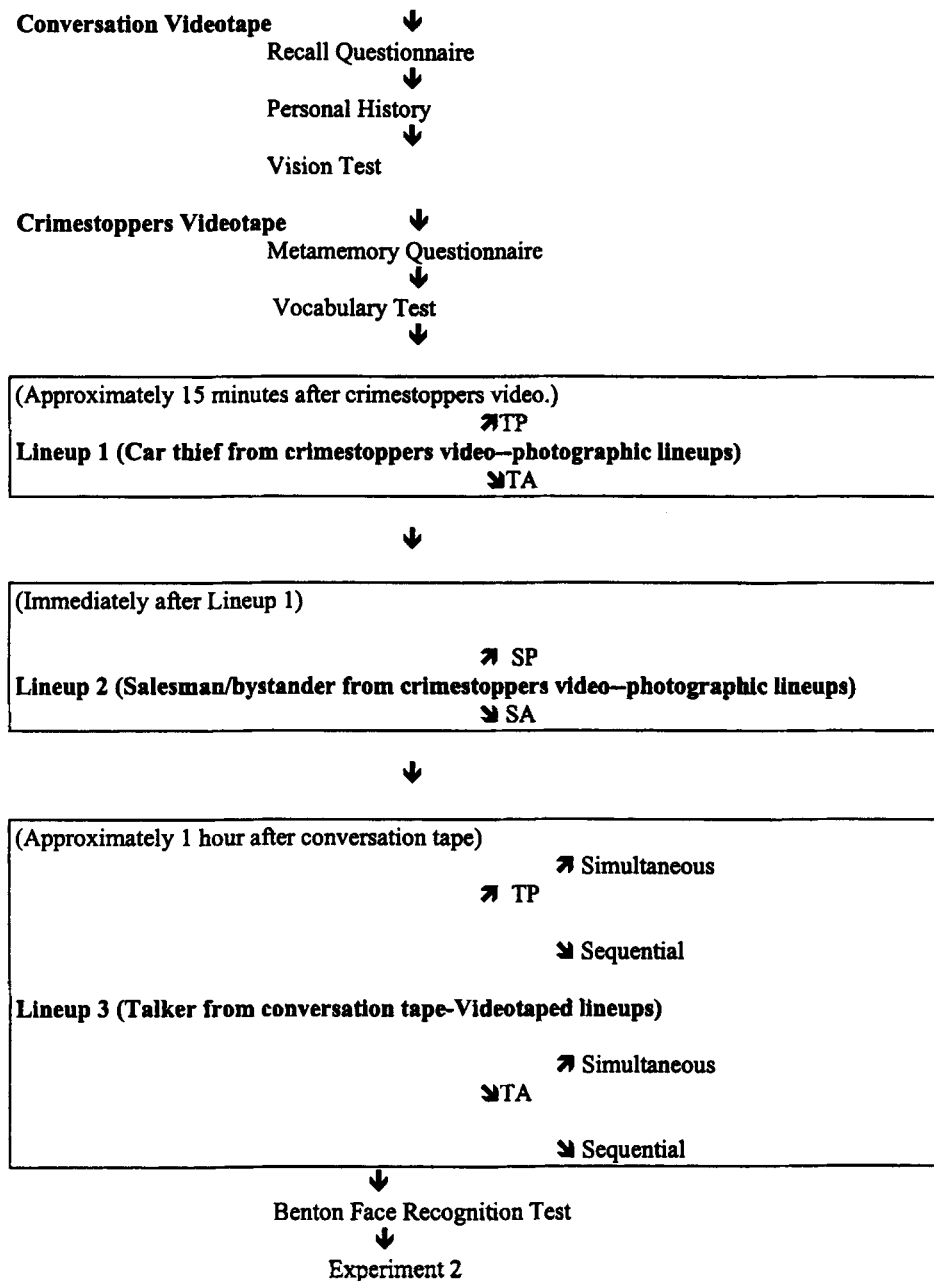
Summary of Design & Procedure

Conversation Videotape   ⬇
Recall Questionnaire
⬇
Personal History
⬇
Vision Test

Crimestoppers Videotape   ⬇
Metamemory Questionnaire
⬇
Vocabulary Test
⬇

(Approximately 15 minutes after crimestoppers video.)
↗TP
Lineup 1 (Car thief from crimestoppers video--photographic lineups)
↘TA

⬇

(Immediately after Lineup 1)

↗ SP
Lineup 2 (Salesman/bystander from crimestoppers video--photographic lineups)
↘ SA

⬇

(Approximately 1 hour after conversation tape)
↗ Simultaneous
↗ TP
↘ Sequential

Lineup 3 (Talker from conversation tape-Videotaped lineups)

↗ Simultaneous
↘TA
↘ Sequential

⬇
Benton Face Recognition Test
⬇
Experiment 2

Figure 1. Summary of sequence of events, design, and procedure. Arrows at angles point to between-groups treatment differences. Vertical arrows indicate sequence of events for all subjects.

fulfillment of a course requirement. The subjects' vision was tested using the Snellen eye chart, and those whose vision in the best eye was less than 20/50 were excluded. Although the mean educational level of the older group (just under 4 years of college) was somewhat higher than that of the younger group (2 years of college), a vocabulary test, consisting of the second half of the vocabulary portion of the Wechsler Adult Intelligence Scale–Revised, showed minimal age differences: Ms = 20 versus 21 for seniors and young adults, respectively (SDs = 9 and 7.5 for seniors and young adults, respectively).

Materials

Videotapes. In the course of the experiment, two different video-tapes were used. One was a full color reenactment of an actual crime (the "crimestoppers tape"), previously produced by a local television station in cooperation with the Dallas Police Department. It had been shown on the air locally once, 7 years prior to our study, and no subject reported having seen it before. The crime depicted was the robbery of a car from a new car dealership. Two men, who were clearly together, entered the dealership; one was taken on a

test drive by the salesman, while the other was seen to remain behind (the car was a two-seater). On the test drive, the "customer" pulled a gun on the salesman, ordered him from the car, and drove away. The tape lasted 1 min and was accompanied by a voice-over describing the events. Although the salesman and the thief were in view during most of the tape, the thief's companion was seen only at the beginning.

The subjects also saw a 2.5-min color videotape depicting a young man discussing several events in his recent experience: a trip to the state fair, a motorcycle race, and a theatrical production (the "conversation tape").

Written questionnaires. All subjects completed a one-page general health and personal information questionnaire, a recall questionnaire based on the conversation tape, and a lineup metamemory questionnaire. The recall questionnaire included 3 questions about the speaker (his age, eye color, and hair color), and 10 questions concerning the events that he described. The metamemory questionnaire included 15 specific questions about the subjects' confidence in their ability to identify the culprit from the crimestoppers tape under various conditions (e.g., "I would be able to identify the car thief in a lineup if I saw him in person"; "I would be able to identify the car thief in a lineup if I saw him in a color photograph").

BFRT. The test was administered as described in Benton et al. (1994). We used the full version of the test and did not adjust scores for age or education. The test consists of 22 trials, in each of which a target face is displayed above six comparison faces. The task is to find from one to three comparison faces that match the identity of the target face. Differences in lighting and pose make the task difficult on some trials.

Lineups. Since false identifications bear most directly on the possible extension of the false alarm effect to the eyewitness paradigm, we employed both target-present and target-absent lineups. Black-and-white head shots measuring 8×10 in. apiece, including photos of the "robber" and the "car salesman," were obtained from a talent agency. They were used to construct two lineups, each consisting of six pictures. The first lineup (Lineup 1) was for the possible identification of the robber. The second (Lineup 2) was ostensibly for the possible identification of the companion who entered the dealership along with the thief. However, the real purpose of the lineup was to investigate bystander effects. The companion was never included in the lineup; instead, either the victim/car salesman or a substitute was used. The subjects also saw a third lineup (Lineup 3) for the identification of the young man in the conversation tape seen at the beginning of the experiment. The lineup again consisted of black-and-white pictures of six different individuals, but in this case they were presented on a 21-in. television screen, in multiple views, rather than as photographs, to permit a comparison of simultaneous versus sequential presentation while controlling the viewing time for each face in each condition (see Procedure). Each picture was about 3×5 in. in size, and the subjects sat about 3.5 ft. from the screen.

Design

Aside from age group, the between-groups variables all involved aspects of lineup presentation. As is shown in Figure 1, all subjects saw three lineups, with Lineups 1 and 2 testing memory for the crimestoppers tape and Lineup 3 testing memory for the conversation tape. Half the subjects saw a version of Lineup 1 that had the robber present (target present, or TP), and half saw the same lineup but with a foil substituted (target absent, or TA). Similarly, half the subjects saw a version of Lineup 2 that had the car salesman in it (salesman present, or SP), and half saw the same lineup but with a foil substituted (salesman absent, or SA). The ostensible target of Lineup 2 was the robber's companion, but the companion did not appear in any lineup. Each subject who saw a TP version of Lineup 1 saw an SA version of Lineup 2, and vice versa. Lineup 3 also had

TP and TA versions. Half the subjects who saw each version of Lineup 3 (TP or TA) had previously seen the TP version of Lineup 1, and half had seen the TA version of Lineup 1. Lineup 3 also included the additional variable of simultaneous versus sequential presentation, which was factorially combined with the TP/TA variable.

Procedure

The subjects were tested individually. Although they were aware that the study concerned eyewitness memory, they were given no specifics regarding forthcoming tests. After filling out questionnaires (see Figure 1), the subjects saw the conversation videotape. They were instructed to imagine that they were sharing a waiting room with the young man and were listening just to pass the time. They performed the tasks listed in Figure 1 before viewing the crimestoppers video, which they were told was a reenactment of a crime. Next came Lineups 1 and 2. After completing Lineup 2, the subjects took a break and continued with the other paper and pencil measures. The last lineup-related task, Lineup 3, was to identify the speaker in the conversation video. Finally, the subjects took the BFRT and proceeded to Experiment 2 (see Figure 1). Details of lineup instructions and presentation are given in the following paragraphs.

Crimestoppers lineups. The subjects were instructed to imagine that the police had a suspect in custody who they thought might be the robber. Witnesses were going to be shown photos to see whether the man could be identified. However, the police could have made a mistake, so that the culprit might not actually be in the lineup. The instructions also stated that the hair and clothing of the suspect might differ from the robber's hair and clothing at the time of the crime, since the photos were taken at a different time.

The six photographs that constituted Lineup 1 were laid out simultaneously on a table in front of the subject, who either indicated a choice or marked a box marked "None of Them" on the response sheet provided. The subject then made a confidence rating on a 1–7 scale, with 7 being *most confident* and 1 being *least confident*. Then the subject was told that a second lineup would be shown to determine whether he or she could identify the robber's companion, who had entered the car dealership with him. In fact, the companion was never present, although for half of the subjects, the salesman was. The instructions and choice procedure were the same as for the previous lineup.

Conversation video lineup. Because we wanted to equalize the total viewing time for each face in each presentation condition (simultaneous and sequential), lineups were shown on video. The same general instructions were given as in previous lineups, appropriately modified for video presentation. The subjects in the simultaneous condition saw a lineup consisting of six views of six faces. First, head-on smiling shots of the lineup were displayed on the screen in two rows of three faces, for 30 sec. Next, left ¾ profiles of each person appeared and were shown for 30 sec, each in the same location as its corresponding previous front view. This was followed by left profiles, right ¾ profiles, right profiles, and finally another frontal shot, this time with a neutral expression. The subjects in the sequential condition saw the same pictures for the same amount of time, but all six pictures of each individual were presented together with one individual following another. In the simultaneous condition, the subjects could make their choices whenever they wished until the end of the last set of views. In the sequential condition, the subjects were required to make a *yes* or *no* response to each person as he was seen; their answer sheets contained spaces for responses to eight individuals, even though only six faces were presented.

Results

We will first present data from the BFRT, and then turn to analyses of accuracy and choosing on Lineups 1

and 3 (i.e., the two lineups with both TP and TA conditions). Analyses pertaining to the bystander effect in Lineup 2 will be presented next, followed by results from the metamemory and person recall measures. The alpha level for all statistical tests was set at .05.

BFRT Scores Between and Within Age Groups

As expected, raw scores (unadjusted for age) on the BFRT showed a main effect of age group $[F(1,150) = 38.8, MS_e = 14.3]$, with scores averaging 47.5 ($SD = 3.23$) among young adults and 43.6 ($SD = 4.28$) among seniors. A subgroup of 17 seniors had raw scores below the "normal" cutoff on the test (41 out of 54). Only one young subject, who appeared to have misunderstood the instructions, scored below 41; he was dropped from subsequent analyses, producing Ns of 76, 58, and 17 for the young, normal BFRT senior, and low BFRT senior groups, respectively. Mean ages in the senior BFRT categories were 70.2 ($SD = 4.72$) for the normal BFRT seniors and 72.8 ($SD = 4.76$) for low BFRT seniors.

Accuracy and Choosing in Lineups 1 and 3

To gain a first perspective on how lineup performance was related to age and BFRT scores, we examined (1) overall accuracy (rates of choosing the target in TP lineups and refraining from choosing in TA lineups), by combining the data from Lineups 1 and 3 (one point given for each accurate response); and (2) overall choosing rates on the same two lineups (one point for each choice, whether accurate or inaccurate). Since there were two lineups, a maximum score of two points was possible on each measure. As is shown in Table 4, accuracy fell from the young group to the normal senior group to the low BFRT senior group, while the choosing rate was higher in both groups of seniors than it was in young adults.

Because of the discrepant cell sizes and high standard deviations, we analyzed the overall accuracy and choosing rates with Kruskal–Wallis analysis of variance (ANOVA). The ANOVA of the accuracy scores showed a main effect for group $[H(2, N = 152) = 21.6]$, and planned comparisons revealed that low BFRT seniors had significantly lower accuracy than did the other seniors $[H(1,N = 75) = 4.90]$, who in turn had lower accuracy than did the young adults $[H(1, N = 135) = 11.03]$. A similar Kruskal–Wallis ANOVA of the choosing rates also supported the group effect $[H(2, N = 152) = 12.53]$, and planned com-

Table 4
Mean Total Accuracy and Total Choosing (and Standard Deviations) on Lineups 1 and 3 for Young Adults, Normal BFRT Seniors, and Low BFRT Seniors

Group	Total Accuracy		Total Choosing	
	M	SD	M	SD
Young	.92	.72	1.24	.65
Normal BFRT senior	.51	.57	1.60	.56
Low BFRT senior	.18	.39	1.59	.71

Note—Both accuracy and choosing scores ranged from 0 to 2 per subject.

Table 5
Proportion Choices (and Standard Deviations) of Targets and Foils in Target-Present (TP) Lineups, and of Target Replacements and Other Foils in Target-Absent (TA) Lineups, Made by Young and Senior Subjects in Lineups 1 and 3 (Low BFRT Seniors Excluded)

	TP Lineups				TA Lineups			
	Target		Foils		Target Replacement		Other Foils	
	Prop.	SD	Prop.	SD	Prop.	SD	Prop.	SD
Lineup 1								
Young	.26	.45	.42	.50	.03	.16	.61	.50
Senior	.19	.40	.63	.49	.00	.00	.81	.40
Difference	.07		−.21		.03		−.20	
Lineup 3								
Young	.68	.47	.05	.23	.00	.00	.42	.50
Senior	.39	.50	.39	.50	.10	.31	.70	.47
Difference	.29		−.34		−.10		−.28	

Note—TP and TA lineups were identical except that the target face in a TP lineup was replaced with an additional foil (the target replacement) in the corresponding TA lineup. The ns for Lineup 1 were 38 and 38 for young adults in the TP and TA conditions, respectively, and 27 and 31 for seniors in the TP and TA conditions, respectively. The corresponding ns for Lineup 2 were 38, 38, 28, and 30.

parisons revealed that the normal BFRT seniors again differed from the young adults $[H(1,N = 135) = 10.63]$, although not from other seniors.

To examine performance in each lineup individually, we performed standard ANOVAs of the data from the young and senior groups with the low BFRT scorers excluded (removing the problem of highly discrepant cell sizes). Table 5 shows the mean rate of correct target choices and erroneous foil choices in the TP conditions of Lineups 1 and 3, along with the mean rate of erroneous choices of (1) the foils that replaced the targets, as well as (2) all other foils, in the TA conditions of these lineups. The "other foils" used in the TA lineups were the same as those used in the TP lineups (i.e., they were the five foils common to the TP and TA lineups). Hence, when comparing the extent of erroneous foil choices across the TP and TA conditions, we will analyze selections of just these five foils.

Table 5 suggests that correct target choices in the TP condition (column 1) were less frequent among seniors than among young adults, although only in Lineup 3. In line with this conclusion, ANOVAs of correct target choices in the TP condition supported a reliable age effect in Lineup 3 $[F(1,62) = 6.14, MS_e = .23]$, although not in Lineup 1. Much more obvious in Table 5 is that older subjects made more erroneous foil choices, regardless of whether the target was present (column 2) or absent (column 4). Indeed, ANOVAs of choices of the five foils common to the TP and TA conditions supported the age main effect in both lineups $[F(1,130) = 6.07, MS_e = .23$, for Lineup 1; $F(1,126) = 18.1, MS_e = .19$, for Lineup 3]. The ANOVA of Lineup 3 also showed a reliable main effect of the TP/TA variable $[F(1,126) = 20.9, MS_e = .19]$, supporting the observation that distractor

Table 6
Proportion Choices (and Standard Deviations) of Salesman and Other Foils in the Salesman-Present (SP) Lineup, and Salesman Replacement and Other Foils in the Salesman-Absent (SA) Lineup, Made by Young and Senior Subjects (Low BFRT Seniors Excluded)

	SP Lineup				SA Lineup			
	Salesman Foil		Other Foils		Salesman Replacement		Other Foils	
	Prop.	SD	Prop.	SD	Prop.	SD	Prop.	SD
Young	.13	.34	.21	.41	.00	.00	.37	.49
Seniors	.06	.25	.58	.50	.00	.00	.44	.51
Difference	.07		−.37		.00		−.07	

Note—The SP and SA lineups were identical except that the salesman's face in the SP lineup was replaced with an additional standard foil (the salesman replacement) in the SA lineup. The ns were 38 and 38 for young adults in the SP and SA conditions, respectively, and 31 and 27 for seniors in the SP and SA conditions, respectively.

choices were less frequent when the target was present than when it was not. However, the age main effect was not qualified by this variable.

Although both ANOVAs of Lineup 3 included the lineup organization (simultaneous vs. sequential) variable, it had no reliable effects. This departs from the usual finding that sequential presentation reduces inaccurate choices without reducing accurate choices (Wells & Seelau, 1995).

Effects of Bystander Presence in Lineup

Table 6 shows performance by age group on Lineup 2, wherein the victim/car salesman appeared in one condition (SP), but not the other (SA). Choices of the victim/salesman in the SP condition occurred only rarely, regardless of age group. However, choices of distractors (excluding the salesman in the SP condition and his replacement in the SA condition) once again were more frequent in the senior group. An ANOVA of distractor choices including age and SP/SA as variables supported the age difference [$F(1,130) = 7.23, MS_e = .23$] and also produced a marginal age × SP/SA interaction [$F(1,130) = 3.14, MS_e = .23; p < .08$]: Whereas seniors made more distractor choices in the SP condition than in the SA condition, the younger subjects showed the opposite trend. We note that the seniors made reliably more choices overall when the salesman was present than when he was not (overall choosing rates of .64 and .44, respectively, a significant difference by Fisher's exact test), whereas the young adults showed no such trend (overall choosing rates of .34 and .37, respectively). This result with seniors might be viewed as a type of bystander effect, since the victim's presence appeared to increase erroneous choosing, even though identifications of the salesman per se occurred very rarely.

Confidence and Metamemory

Mean ratings of confidence across the three lineups showed no reliable age difference, averaging 4.1 ($SD = 1.1$) for young adults and 4.0 ($SD = 1.5$) for seniors on the 1–7 scale (7 = *highest confidence*). However, older subjects rated themselves lower in memory self-efficacy than did the younger adults on the prelineup questionnaire; this difference was significant, though not dramatic. There were two measures: the first was the number of lineup conditions (out of 15) under which subjects believed they could identify the culprit. The mean for seniors was 9.7 ($SD = 4.16$), and for the young, 11.7 ($SD = 2.29$) [$F(1,132) = 13.8, MS_e = 11.2$]. The second measure was the average confidence level for "yes" responses on the questionnaire, excluding confidence ratings of less than 20% sure. Here, the mean for the seniors was .63 ($SD = .24$) and for the young was .71 ($SD = .15$), a difference that was also reliable [$F(1,132) = 5.71, MS_e = 397.8$].

To assess the relation between prospective confidence (on the self-efficacy questionnaire) and retrospective confidence (on the lineups), we normalized the two self-efficacy scores, averaged the z scores to get a single prospective self-efficacy score, and then computed the gamma correlations. Both age groups showed significant correlations between prelineup estimates of the ability to recognize the target in Lineup 1 and postlineup confidence that they had responded correctly ($\gamma = .19$ and .33 for young and elderly, respectively). The seniors showed significant correlations between their metamemory predictions concerning Lineup 1 and confidence on the other two lineups as well ($\gamma = .25$ and .19 for Lineups 2 and 3, respectively). Also, for seniors only, all three postlineup confidence ratings were correlated (all three $\gamma s = .27$). Unfortunately, the reduced correlations in the young adult group are difficult to interpret, because of reduced standard deviations in the confidence measures in this group.

Person Recall

Finally, using the data from the recall questionnaire that subjects completed after hearing the conversation videotape, we counted the number of correct responses (out of 3) to questions about the speaker's appearance. Mean recall scores averaged .62 ($SD = .73$) and .83 ($SD = .65$) for the seniors and young adults, respectively, which was not a reliable difference. To assess the relationship of recall memory to recall accuracy, we computed gamma correlations between the two recall measures and identification accuracy, choosing, and confidence on the lineup in which the speaker was the target (Lineup 3), as well as overall choosing and accuracy rates. Neither recall measure correlated significantly with any lineup measure (largest $\gamma < .12$) in either age group.

Discussion

Experiment 1 showed that age-related deficits in face recognition, and specifically the false alarm effect, generalize to the context of eyewitness identification. The key finding of an age-related increase in erroneous lineup choices was found in three different lineup tasks that differed in difficulty, delay, and presentation mode.

Further research, perhaps incorporating more ecologically valid parameters, will be needed to establish the generality of the false alarm effect within the domain of eyewitness identification. The present research, though just a first step, serves to raise the possibility that the criminal justice system may need to reform procedures to take into account the increased probability of false identifications by older adult witnesses.

Although seniors exceeded younger subjects in making false identifications in Lineups 1 and 3, they were not more prone to select the "bystander" (i.e., the salesman) in Lineup 2. This outcome presents a problem for the context recollection hypothesis, which had predicted an increase in bystander selections among the older subjects. However, hope for the hypothesis is provided by the finding that older subjects made generally more choices in the SP condition than in the SA condition, whereas the young adult subjects did not. It is possible that the presence of the salesman, who had been seen earlier, inflated seniors' sense of perceived familiarity of the lineup as a whole. Although they seldom were fooled into choosing the salesman (perhaps reflecting some ability to recollect context), their heightened sense of general familiarity might have increased their already significant bias to make a lineup choice. This post hoc interpretation will be addressed in future studies.

What about alternative accounts of increased choosing rates among elderly eyewitnesses? One such alternative is that our seniors simply failed to understand, or did not believe, that targets truly might be absent from our lineups. More research may be necessary for this hypothesis definitely to be rejected, but we note that our lineup instructions were designed to discourage unreflective choosing by elaborating and repeating the point that the police might be wrong, that the true culprit might not be in the lineup, and that it was not necessary to make a choice. Also, the no-choice option was prominently displayed on the response sheets (which also included another reminder that the target might not be present). It would seem to be have been difficult for seniors to miss all these reminders.

Another view of our findings is that perceptual processing deficits were the cause of seniors' high choosing rates. We included the BFRT specifically in order to explore this possibility. We found that while low BFRT scores were predictors of low accuracy, they were *not* predictors of high choosing rates. Moreover, age-related increases in choosing rates were large and reliable even when low scorers on the BFRT test were dropped from the elderly sample. Although other types of perceptual deficits besides those tapped by the BFRT might be confounded with age and affect lineup choosing, we think this unlikely. The BFRT closely resembles the lineup task in the demands that it makes on perceptual processes (one person's face must be picked out from others despite changes in pose, expression, etc.). Hence, perceptual deficits not tapped by the BFRT are unlikely to influence lineup identification.

It also is important to consider the possibility that seniors' generally poorer memories are sufficient to explain their elevated choosing rates. Weighing against this interpretation, we found that poor recall of information from the conversation video was not associated with higher rates of false choosing, or with generally lower accuracy, in either age group. This suggests that no general "poor memory" factor was operative. In a subsequent study (Searcy, Bartlett, & Memon, 1998), we found that person recall correlated *positively* with erroneous choosing, particularly among seniors. These findings make it even more unlikely that generally poor memory mediates the false alarm effect. Indeed, it would appear that good recall, if anything, encourages, rather than discourages, false identification.

Although we doubt that failures to comprehend instructions, perceptual processing deficits, or generally poor memory can explain the elevated false choosing rates among our senior subjects, the context recollection hypothesis is not uniquely capable of handing the results. In our introduction, we argued that the literature on old/new recognition suggests two alternative views—that older persons have problems in making fine mnemonic discriminations, and that they suffer a general deficit in visuospatial memory. Either view can handle the present observations of age-related increases in lineup choosing rates; indeed, it is arguable that either idea would have predicted such differences. In Experiment 2, we returned to the old/new recognition paradigm and addressed a more distinctive and counterintuitive prediction of the context recollection hypothesis.

EXPERIMENT 2
False Alarms for Conjunctions and New Faces

Recent research in old/new recognition (Reinitz, Lammers, & Cochran, 1992; Reinitz, Morrissey, & Demb, 1994) has examined false alarm rates for a type of lure item that combines the features of two different study items within a perceptually well-integrated stimulus. False alarm rates for such *conjunction* lures have been examined in various stimulus domains, including words, (nonsense) pseudo-words, and more and less elaborated line drawings of faces (Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996; Reinitz et al., 1992; Reinitz et al., 1994). Invariably, such lures evoke high rates of false alarms—reliably higher than entirely new items or *feature lures* composed of half old and half new features.

A number of considerations, including common sense, suggest that older individuals should be particularly susceptible to false recognitions of conjunction lures. First, since age-related deficits generally are present in more difficult tasks, one might predict on this basis alone that older persons would be worse than young adults in rejecting conjunction lures. Moreover, false alarm rates for conjunctions appeared to be higher among healthy older persons than among young adults in the Kroll et al. (1996) study (see Table 3), although age differences were not

statistically evaluated there. Finally, conjunction foils have two characteristics that we have identified as being associated with age-related false alarm rate differences: They closely resemble study faces, and their rejection as "new" would appear to require encoding and retrieval of visuospatial information. Hence, if there were age-related deficits in either fine-discrimination ability or in visuospatial memory, we would predict age differences with these items.

Such considerations notwithstanding, the context recollection hypothesis carries the implication that false alarm rates for conjunctions might *not* show age differences under some conditions, perhaps especially conditions of face recognition. This implication derives from Mandler's (1980) conception that perceived familiarity is based on the intra-item integration or organization of the sensory and perceptual elements of a stimulus, apart from any memory for the context(s) of its prior encounter. A conjunction clearly differs from its *parent* stimuli in terms of intra-item relations, and, therefore, it should be distinguishable from an old face on the basis of familiarity. If perceived familiarity is largely age invariant (Bartlett, 1993; and see Dywan & Jacoby, 1990; Jacoby et al., 1996; Rabinowitz, 1984), false alarm rates for conjunctions might show minimal age differences.

In applying Mandler's (1980) theory to the conjunction effect, we would not wish to claim that old-conjunction discrimination is based only on differences in perceived familiarity under all conditions. More complex retrieval processes involving context recollection might contribute as well, especially in conditions that impair the encoding of intra-item relations (e.g., the use of stimuli that are not perceived as wholes). With these considerations in mind, we attempted to facilitate the encoding of intra-item relations by (1) using photographic quality pictures of faces as our stimuli, (2) using image graphics techniques to construct "seamless" composite faces, and (3) presenting each face picture two times at study. A study list including 16 faces, each presented two times, was followed by a test containing 8 copies of study list faces (old items), 8 conjunctions, each of which included the internal features of one study face and the external features of another, and 8 entirely new faces. We expected to replicate the finding (Table 1) of an age-related increase in false alarm rates for entirely new faces. The key question was whether we would also observe an age-related increase in false alarm rates for conjunctions.

Method

Subjects

Experiment 2 was conducted at the conclusion of Experiment 1, using the same subjects. However, 2 young adults were dropped because they had seen the materials before. This left 74 young adults, 58 normal BFRT elderly, and 17 low BFRT elderly.

Materials

The stimuli were black-and-white photographs of male faces from a 1965 Texas A&M yearbook. The original images were digitized and the composites created with Corel Draw software. "Noise" was added to all of the images to regularize the skin texture across pictures, which facilitated the creation of realistic composites. Composites consisted of the inner features of one face (eyes, nose mouth) and the outer features of another (hair, ears, chin). The contiguous edges of the two components of each composite were blended with Corel Draw, creating faces that were superficially indistinguishable from original yearbook photographs. Examples of two originals and a composite made from them are shown in Figure 2.

Procedure

The faces were approximately 6 × 6 in. in size when projected on the 21-in. monitor, and they were viewed from a distance of 3.5 ft. Viewing time at study and test was 5 sec per item, with a 3-sec ISI during which the serial position of the next face was displayed. The subjects were instructed that a set of faces would be shown and that a recognition test would follow. The list itself included 16 faces, each presented two times, along with 3 filler faces. The 35 items were randomly sequenced, with the constraints that the fillers occupy the first and last two serial positions, and that each of the 16 critical faces appear once in each list half. Half of the critical 16 study faces were original photographs, and half were composites constructed from two different original photographs as described above.

Following the study list, the subjects saw a 24-item test list that consisted of (1) 8 items from the study list, (2) 8 items that were conjunctions of faces from the study list, and (3) 8 entirely new items. Half of the faces in each condition were original yearbook photographs, and half were composites that combined the inner and outer features of two different yearbook photographs.
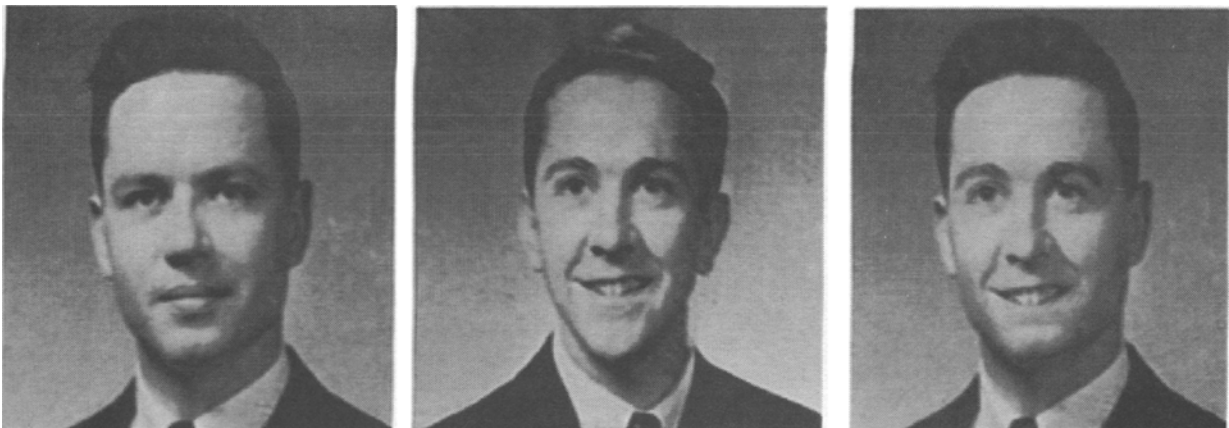

Figure 2. Examples of two original faces and one composite of the type used in Experiment 2.

Table 7
d's, Hit Rates, and False Alarm Rates for Conjunctions (FAC) and New Items (FAN) on the Face Recognition Test (With Standard Deviations) for Young Adults, Normal BFRT Seniors, and Low BFRT Seniors

							d'
	Hit Rates		FAC		FAN		Old/New		Old/Conj.	
Age Group	M	SD	M	SD	M	SD	M	SD	M	SD
Young	.79	.18	.39	.22	.06	.11	2.93	1.17	1.40	.92
Seniors										
Normal BFRT	.80	.21	.40	.20	.20	.20	2.30	1.43	1.45	1.13
Low BFRT	.69	.18	.53	.24	.30	.24	1.17	.63	.45	.95

Note—The large ns of 76 and 58 for the young and normal BFRT seniors produce standard errors for the hit, FAC, and FAN scores of less than .03 in these two groups. The smaller n of 17 for the low BFRT seniors puts their standard errors at less than .06.

Hence, half of the conjunctions on the recognition test were in fact original yearbook photographs whose parts had been previously viewed in two different composites in the study list. The remaining conjunctions were composites of yearbook photographs whose parts had been previously viewed in two different originals in the study list.

Results

Table 7 shows performance on the old/new face recognition task by the young adults and by the older subjects who scored normally and poorly on the BFRT. Note that the hit rates for old faces were virtually identical in the young and normal BFRT senior groups, and only slightly reduced in the low BFRT senior group. Indeed, an ANOVA of the hit rates showed no group effect ($p > .10$). In contrast, the false alarm rates for conjunctions (FAC) and entirely new faces (FAN) showed an interactive pattern whereby (1) the normal BFRT elderly exceeded the young in FAN but not in FAC and (2) the low BFRT elderly exceeded the normal BFRT elderly in both types of false alarm. Supporting this characterization, an ANOVA performed on the false alarm rates showed main effects of both group [$F(1,146) = 11.9, MS_e = 444$] and lure type [$F(1,146) = 110.5, MS_e = 295$], as well as the interaction [$F(2,146) = 4.73, MS_e = 295$]. Moreover, all three effects remained when the low BFRT seniors were dropped from the ANOVA, leaving only the young and normal BFRT groups [$Fs(1,130) = 8.47, 160.5$, and 9.44, for group, lure type, and the interaction, respectively, with MS_es = 405, 285, and 285, respectively]. A separate ANOVA comparing just the normal and low BFRT elderly showed only the group and lure type main effects ($Fs = 5.57$ and 44.4, respectively, with MS_es = 604 and 281, respectively), with no interaction.

The pattern of hits and false alarms suggests that the normal BFRT seniors differed from young adults in the relatively easy discrimination between old and entirely new faces, but not in the more difficult discrimination between old and conjunction faces. To evaluate this implication, we used the hit and false alarm rates of each subject to compute two d' scores, one for old/new discrimination and another for old/conjunction discrimination. As shown in Table 7, differences between young adult

subjects and normal BFRT seniors were limited to the old/new scores. Planned comparisons revealed that the young exceeded the normal BFRT seniors in old/new d's [$F(1,149) = 12.0, MS_e = .54$], but not in old/conjunction d's ($F < 1$). In contrast, the low BFRT seniors fell below the normal BFRT seniors in old/conjunction d's [$F(1,149) = 4.44, MS_e = .75$], though not significantly in old/new d's ($F < 1$).[3]

Discussion

Young adults and normal BFRT seniors showed virtually identical hit rates, and they showed no difference in false alarm rates for conjunctions of old faces. However, in agreement with previously published face recognition studies, as well as with the findings from Experiment 1, false alarm rates for entirely new faces showed an age-related increase. The d' analysis reinforces the data from hits and false alarms. The usual age-related deficit in d's was supported, but only when the index was based on data from new items. The d's for discriminating old items from conjunctions showed no deficit for normal BFRT seniors in comparison with young adults. Although the old/conjunction d's were relatively low, they were not insensitive, as was revealed by the difference between the normal BFRT and low BFRT groups.

The outcome of Experiment 2 strongly supports the context recollection hypothesis outlined in the introduction. The hypothesis holds that false alarms to new faces are increased in old age because the age-sensitive process of context recollection aids successful rejection of such faces. False alarm rates for conjunctions are not increased in old age, because such items are more easily rejected on the basis of differences in perceived familiarity—itself based on intra-item integrative and relational information—than on the basis of more cognitively complex processes involved in context recollection. Older persons free of serious perceptual deficits generally show no impairments in the processing activities underlying the experience of perceived familiarity. This is the reason why they showed no deficit in rejecting conjunction lures.

We would not make this argument for all types of conjunctions in all experimental situations. Faces are un-

usual—if not actually unique—in the extent to which they are processed as well-integrated wholes (e.g., Bartlett & Searcy, 1993; Bruce & Humphreys, 1994; Tanaka & Farah, 1993). With stimuli less conducive to holistic processing, the integrative mechanisms underlying the experience of perceived familiarity might not support discrimination between old items and conjunctions. In such cases, there might be no other basis than conscious recollection for old/conjunction discrimination. Indeed, such reasoning might explain the appearance of age differences in false alarm rates for conjunctions in prior studies using stimuli such as actions (Cohen & Faulkner, 1989) or highly schematized face drawings (Kroll et al., 1996), summarized in Table 3.

Future research must address the range of stimuli and tasks that can produce the present pattern of an age-related increase in false alarm rates for new items along with age invariance in false alarm rates for conjunctions. However, regardless of what such research shows, the present findings are theoretically important in at least two ways. First, they are important as a counterexample to the general observation that more difficult tasks produce larger age deficits, and as a new bit of evidence that age-related increases in false alarm errors do not simply reflect generally poorer perceptual or mnemonic processing on the part of older persons. Normal BFRT seniors performed as well as young subjects on the difficult task of rejecting conjunction lures. These seniors were impaired in comparison to the young *only* on the less difficult task of rejecting entirely new lures. If generally poor perception or memory were the causal factor producing the false alarm effect, it would seem to follow that age differences in false alarm rates would appear across the board or would be even larger with difficult lures than with easier lures. Yet only the low BFRT seniors showed a general performance deficit; the deficit shown by the normal BFRT seniors was much more specific.

The second way in which our findings are important is that they violate predictions made by two otherwise quite plausible alternative hypotheses considered in the introduction. One of these hypotheses—consistent with a large number of findings summarized in Table 3—holds that older individuals suffer deficits in making fine discriminations between old stimuli and highly resembling lures. The second hypothesis—supported by the finding of false alarm effects with novel, visuospatial stimuli—holds that older individuals are generally impaired in visuospatial memory. Conjunctions clearly qualify as highly resembling lures, and distinguishing such items from old faces would appear to be a task involving visuospatial memory. Hence, both hypotheses should predict the occurrence of strong age-related differences in rejecting conjunction lures. In violating this prediction, the present findings rule against these two hypotheses as accounts for age differences in face recognition. Of course, they remain as viable explanations for age differences in other stimulus domains (see Koutstaal & Schacter, 1997).

GENERAL DISCUSSION

The results of our experiments provide important information about the generality and processing mechanisms of the false alarm effect in face recognition. In Experiment 1, we confirmed that the false alarm effect generalizes to the eyewitness identification paradigm. While accurate choices in TP lineups showed inconsistent age differences, false identification rates were consistently higher among seniors (see Tables 4 and 5). That the false alarm effect found in old/new recognition would extend to the lineup tasks was predicted by a context recollection hypothesis—which holds that age-related increases in false alarm errors reflect age-related deficits in recollecting context—but it also was consistent with two alternative hypotheses considered in the introduction. In Experiment 2, we tested the more discriminating prediction that age-related increases in false alarm errors might not generalize from entirely new lures to conjunction lures that combine the features of different study faces. This counterintuitive prediction was impressively confirmed, favoring the context recollection hypothesis over both of the alternatives that we had been considering. In further support of the context recollection hypothesis, both experiments ruled against the possibility that the false alarm effect in face recognition is due to a deficit in perceptual processing of faces. Although seniors who scored poorly on a test of face matching (the BFRT) showed performance impairments, these differed qualitatively from those shown by other seniors. Moreover, the false alarm effect in old/ new recognition and lineup identification remained very much in evidence when the seniors showing evidence of perceptual impairments were excluded from the analyses.

Although the present data favor the context recollection hypothesis, this and other new research suggests that it might require modification or elaboration. The context recollection hypothesis as formulated here suggests that the old rely *more* on familiarity than the young do. An alternative construal (see Bartlett, Halpern, & Dowling, 1995) holds that perceived familiarity affects recognition regardless of age. However, attaining optimal levels of recognition accuracy often requires that observers distinguish different *sources* of perceived familiarity (see also Schacter, Koutstaal, Johnson, Gross, & Angell, 1997). Recollection of contextual information provides one way to solve the source discrimination problem. But since the older adults are impaired at context recollection, they will tend to show impairments in recognition accuracy, and specifically in rejecting highly familiar lures, when (1) source discrimination is required by the task, or at least is facilitative to performance on the task, and (2) alternative methods of source discrimination are either unavailable or not very effective. Face recognition may involve both of these factors, since even entirely new faces in a recognition test—or in a lineup—might evoke strong feelings of perceived familiarity (Bartlett, 1993). Moreover, the source of this perceived familiarity may be hard

to determine in any other way than through recall of prior context.

Two important issues remain to be addressed. First, the simultaneous/sequential manipulation in Experiment 1 (Lineup 3) did not confirm prior findings of a reliable reduction in erroneous lineup choosing with sequential presentation (Cutler & Penrod, 1995; Wells, 1984). Although we were surprised by this finding, we note that the simultaneous/sequential variable was not introduced until Lineup 3, by which time all subjects had already experienced two simultaneous lineups. It is possible that the subjects' strategies were too firmly set to be shifted at this stage. Supporting this idea, Parker and Ryan (1993) found that the benefits of sequential presentation were reduced by practice. Alternatively, our unusual method of video presentation may have reduced the effectiveness of the sequential lineup. In a subsequent study (Searcy et al., 1998), we have found significant reductions in choosing by both young adults and seniors in a more standard sequential presentation condition.

A final issue to address concerns the extent to which our lineup study has real-world implications. Several aspects of the eyewitness simulation scenario have been criticized, among them the use of uninvolved bystander witnesses and videotaped presentation of simulated events, both of which were characteristics of Experiment 1. However, the available evidence suggests that laboratory studies yield relevant data. First, uninvolved bystander witnesses are not rare. Lieppe (1995) estimates that approximately 8,000 trials a year may be based on evidence from unaffected bystanders. While this may be a small percentage of all trials, the figure is not negligible. Second, bystanders and victim witnesses do not differ in identification rates (Hosch & Cooper, 1982). Finally, Yuille and colleagues (e.g., Tollestrup, Turtle, & Yuille, 1994; Yuille & Cutshall, 1989) have conducted field and archival analyses of real-life cases, the results of which are mostly consistent with what laboratory studies would predict. Summarizing these data, Lieppe (1995) states that:

> There is practically no evidence of what could be called realism effects ... on memory reports and identifications. There is no documentation of a relationship that has been well established in laboratory paradigms reversing in direction in a more involving and realistic eyewitness setting. (p. 919)

In summary, an age-related increase in false alarm errors is a robust phenomenon in face recognition, with implications for theory as well as for the practice of interrogating older eyewitnesses to crime. The context recollection approach offers promise as an avenue for exploring these implications. It makes the testable prediction that age-related false alarm effects should appear in recognition tasks whenever perceived familiarity is a nonoptimal basis for distinguishing targets from lures and context recollection is helpful. It also predicts that interventions that improve context recall (e.g., training in such recall; see Jacoby et al., 1996), might help to reduce the influence of context-free familiarity information on seniors' real-world recognition and identification decisions.

REFERENCES

Adams-Price, C. (1992). Eyewitness memory and aging: Predictions of accuracy in recall and person recognition. *Psychology & Aging, 7,* 602-608.

Bäckman, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition, 19,* 63-71.

Bartlett, J. C. (1993). Limits on losses in face recognition. In J. Cerella, J. Rybash, W. Hoyer, & M. L. Commons, *Adult information processing: Limits on loss* (pp. 351-379). New York: Academic Press.

Bartlett, J. C., & Fulton, A. (1991). Familiarity and recognition of faces in old age. *Memory & Cognition, 19,* 229-238.

Bartlett, J. C., Halpern, A. R., & Dowling, W. J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition, 23,* 531-546.

Bartlett, J. C., & Leslie, J. E. (1986). Aging and memory for faces versus single views of faces. *Memory & Cognition, 14,* 371-381.

Bartlett, J. C., Leslie, J. E., Tubbs, A., & Fulton, A. (1989). Aging and memory for pictures of faces. *Psychology & Aging, 4,* 276-283.

Bartlett, J. C., & Searcy, J. H. (1993). Inversion and configuration of faces. *Cognitive Psychology, 25,* 281-316.

Bartlett, J. C., Strater, L., & Fulton, A. (1991). False recency and false fame of faces in young adulthood and old age. *Memory & Cognition, 19,* 177-188.

Bartlett, J. C., Till, R. E., Gernsbacher, M. A., & Gorman, W. (1983). Age-related differences in memory for lateral orientation of pictures. *Journal of Gerontology, 4,* 439-446.

Benton, A. L., Eslinger, P. J., & Damasio, A. R. (1981). Normative observations on neuropsychological test performance in old age. *Journal of Clinical Neuropsychology, 3,* 33-42.

Benton, A. L., Sivan, A., Hamsher, K. de S., Varney, N. R., & Spreen, O. (1994). *Facial recognition: Stimulus and multiple choice pictures: Contributions to neuropsychological assessment* (pp. 35-52). New York: Oxford University Press.

Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and validity of the Memory Self-Efficacy Questionnaire. *Developmental Psychology, 25,* 701-713.

Bornstein, B. H. (1995). Memory processes in elderly eyewitnesses: What we know and what we don't know. *Behavioral Sciences & the Law, 13,* 349-363.

Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. *Visual Cognition, 1,* 141-180.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24,* 403-416.

Cohen, G., & Faulkner, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology & Aging, 4,* 10-17.

Craik, F. I. M., & Jennings, J. (1992). Human memory. In F. I. M. Craik & T. Salthouse, *The handbook of aging and cognition* (pp. 51-110). Hillsdale, NJ: Erlbaum.

Crook, T. H., & Larrabee, G. J. (1992). Changes in facial recognition memory across the adult life span. *Journal of Gerontology: Psychological Sciences, 47,* P138-P141.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology and the law.* New York: Cambridge University Press.

Diamond, R., & Carey, S. (1986). Why faces are and aren't special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Dywan, J., & Jacoby, L. (1990). Effects of aging on source monitoring: differences in susceptibility to false fame. *Psychology & Aging, 5,* 379-397.

Ferris, S. H., Crook, C., Clark, E., McCarthy, M., & Raye, C. (1980). Facial recognition deficits in normal aging and dementia. *Journal of Gerontology, 5,* 707-714.

Flicker, C., Ferris, S., Crook, T., & Bartus, R. (1990). Age differences

in the vulnerability of facial recognition memory to proactive inter-ference. *Experimental Aging Research*, **15**, 189-194.

FULTON, A., & BARTLETT, J. C. (1991). Young and old faces in young and old heads: The factor of age in face recognition. *Psychology & Aging*, **6**, 623-630.

GRADY, C. L., McINTOSH, A., HORWITZ, B., UNGERLEIDER, L., MAISOG, J. M., MENTIS, M. J., PIETRINI, P., WAGNER, E., & HAXBY, J. V. (1995). Age-related reductions in human recognition memory due to impaired encoding. *Science*, **269**, 218-221.

HASHTROUDI, S., JOHNSON, M. K., VNEK, N., & FERGUSON, S. A. (1994). Aging and the effects of affective and factual focus on source monitoring and recall. *Psychology & Aging*, **9**, 160-170.

HOSCH, H. (1994). Individual differences in personality and eyewitness identification. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 328-347). New York: Cambridge University Press.

HOSCH, H., BOTHWELL, R. K., SPORER, S. L., & SAUCEDO, C. (1990, April). *The accuracy of a witness' choice: Facial recognition ability, confidence, decision-time, and person descriptions*. Paper presented at the annual meeting of the Southwestern Psychological Associa-tion.

HOSCH, H., & COOPER, D. S. (1982). Victimization as a determinant of eyewitness accuracy. *Journal of Applied Psychology*, **67**, 649-652.

JACOBY, L. L., JENNINGS, J. M., & HAY, J. F. (1996). Dissociating auto-matic and consciously controlled processes: Implications for diagno-sis and rehabilitation of memory deficits. In D. Herrman, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Theory in context* (pp. 161-191). Mahwah, NJ: Erl-baum.

JENNINGS, J. M., & JACOBY, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repe-tition. *Psychology & Aging*, **12**, 352-361.

KAUSLER, D. H. (1994). *Learning and memory in normal aging*. San Diego: Academic Press.

KORIAT, A., BEN-ZUR, H., & SHEFFER, D. (1988). Telling the same story twice: Output monitoring and age. *Journal of Memory & Language*, **27**, 23-39.

KOUTSTAAL, W., & SCHACTER, D. L. (1997). Gist-based false recogni-tion of pictures in older and younger adults. *Journal of Memory & Lan-guage*, **37**, 555-583.

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

LIEPPE, M. R. (1995). The case for expert testimony about eyewitness memory. *Public Policy & Law*, **1**, 909-959.

LIGHT, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, **42**, 333-376.

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

MASON, S. E. (1986). Age and gender as factors in facial recognition and identification. *Experimental Aging Research*, **12**, 151-154.

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.

O'ROURKE, T. E., PENROD, S. D., CUTLER, B. L., & STUVE, T. E. (1989). The external validity of eyewitness identification research: General-izing across subject populations. *Law & Human Behavior*, **13**, 385-395.

PARK, D. C., PUGLISI, J. T., & SMITH, A. G. (1986). Memory for pictures: Does an age-related decline exist? *Psychology & Aging*, **1**, 11-17.

PARK, D. C., PUGLISI, J. T., & SOVACOOL, M. (1984). Picture memory in older adults: Effects of contextual detail at encoding and retrieval. *Journal of Gerontology*, **39**, 213-215.

PARKER, J. F., & RYAN, V. (1993). An attempt to reduce guessing behav-ior in children's and adults' eyewitness identifications. *Law & Human Behavior*, **17**, 11-26.

PEZDEK, K. (1987). Memory for pictures: A life-span study of the role of visual details. *Child Development*, **58**, 807-815.

RABINOWITZ, J. C. (1984). Aging and recognition failure. *Journal of Gerontology*, **39**, 65-71.

RANKIN, J. L., & KAUSLER, D. H. (1979). Adult age differences in false recognitions. *Journal of Gerontology*, **34**, 58-65.

READ, J. D. (1994). Understanding bystander misidentifications: The role of familiarity and contextual knowledge. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 56-79). New York: Cambridge University Press.

READ, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual in-formation. *Applied Cognitive Psychology*, **9**, 91-121.

REDER, L. M., WIBLE, C., & MARTIN, J. (1986). Differential memory changes with age: Exact retrieval versus plausible inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 72-81.

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

SCHACTER, D. L., COOPER, L. A., & VALDISERRI, M. (1992). Implicit and explicit memory for novel visual objects in older and younger adults. *Psychology & Aging*, **7**, 299-308.

SCHACTER, D. L., KOUTSTAAL, W., JOHNSON, M. K., GROSS, M. S., & ANGELL, K. E. (1997). False recollection induced by old photographs: A comparison of older and younger adults. *Psychology & Aging*, **12**, 203-215.

SCHACTER, D. L., KOUTSTAAL, W., & NORMAN, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences*, **1**, 229-236.

SCOGIN, F., CALHOON, S., & D'ERRICO, M. (1994). Eyewitness confidence and accuracy among three age cohorts. *Journal of Applied Gerontol-ogy*, **13**, 172-184.

SEARCY, J. H., BARTLETT, J. C., & MEMON, A. (1998). *Relationship of post-event narratives, lineup conditions and individual differences to false identification by young and older eyewitnesses*. Manuscript submitted for publication.

SMITH, A. D. (1975). Partial learning and recognition memory in the aged. *International Journal of Aging & Human Development*, **6**, 359-365.

SMITH, A. D., PARK, D. C., CHERRY, K., & BERKOVSKY, K. (1990). Age differences in memory for concrete and abstract pictures. *Journal of Gerontology: Psychological Sciences*, **45**, 205-209.

SMITH, A. D., & WINOGRAD, E. (1978). Adult age differences in re-membering faces. *Developmental Psychology*, **14**, 443-444.

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

TILL, R. E., BARTLETT, J. C., & DOYLE, A. (1982). Age differences in picture memory with resemblance and discrimination tasks. *Experi-mental Aging Research*, **8**, 179-184.

TOLLESTRUP, P. A., TURTLE, J. W., & YUILLE, J. C. (1994). Actual vic-tims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). New York: Cam-bridge University Press.

TRAHAN, D. E., LARRABEE, G. L., & LEVIN, H. S. (1986). Age-related differences in recognition memory for pictures. *Experimental Aging Research*, **12**, 147-150.

TRAHAN, D. E., LARRABEE, G. L., & QUINTANA, J. W. (1990). Visual recognition memory in normal adults and patients with unilateral vas-cular lesions. *Journal of Clinical & Experimental Neuropsychology*, **12**, 857-872.

WELLS, G. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, **14**, 89-103.

WELLS, G., & SEELAU, E. (1995). Eyewitness identification: Psycho-logical research and legal policy on lineups. *Psychology, Public Pol-icy & Law*, **1**, 765-791.

YARMEY, A. D. (1984). Age as a factor in eyewitness testimony. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological per-spectives* (pp. 142-154). New York: Cambridge University Press.

YARMEY, A. D. (1996). The elderly witness. In S. L. Sporer, R. S. Mal-pass, & G. Koehnken (Eds.), *Psychological issues in eyewitness iden-tification* (pp. 259-278). Mahwah, NJ: Erlbaum.

YARMEY, A. D., & KENT, J. (1980). Eyewitness identification by elderly and young adults. *Law & Human Behavior*, 4, 359-371.

YUILLE, J., & CUTSHALL, J. (1989). Analysis of statements of victims, witnesses and suspects. In J. Yuille (Ed.), *Credibility assessments* (pp. 175-191). Dordrecht: Kluwer.

NOTES

1. In terms of signal detection theory, the pattern of inconsistent but generally age invariant hit rates and an age-related increase in false alarm rates suggests both a deficit in discrimination accuracy and a lowering (loosening) of criterion in old age. Indeed, the studies that have reported the relevant measures have shown reliable deficits in discrimination and at least trends for criterion reductions in their senior groups. Computing d's and C scores from the hit and false alarm rates for each experiment listed in Table 1, we found that young adults performed with a mean d' of 1.8 and a neutral (unbiased) criterion of $-.03$, whereas seniors performed with a lower mean d' ($M = 1.2$), and a substantially more liberal criterion of $-.30$ [t tests performed on the d's and Cs computed from the 12 studies supported the significance of both effects, ts(11) = 5.41 and 4.93, respectively].

2. This effect was especially strong with the possible (spatially coherent) drawings that might be viewed as most similar to faces. With these drawings, false alarm rates for the young and elderly groups averaged .20 and .46, respectively, with the corresponding hit rates averaging .68 and .75.

3. We also computed bias scores (Cs), which showed that the criterion for old/new discrimination was lower in the two senior groups (Ms = .00 and $-.02$, for the normal and low BFRT seniors, respectively) than in the young adult group ($M = .43$).