

Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease

JAMES C. BARTLETT

University of Texas at Dallas, Richardson, Texas

ANDREA R. HALPERN

Bucknell University, Lewisburg, Pennsylvania

and

W. JAY DOWLING

University of Texas at Dallas, Richardson, Texas

We tested normal young and elderly adults and elderly Alzheimer's disease (AD) patients on recognition memory for tunes. In Experiment 1, AD patients and age-matched controls received a study list and an old/new recognition test of highly familiar, traditional tunes, followed by a study list and test of novel tunes. The controls performed better than did the AD patients. The controls showed the "mirror effect" of increased hits and reduced false alarms for traditional versus novel tunes, whereas the patients false-alarmed as often to traditional tunes as to novel tunes. Experiment 2 compared young adults and healthy elderly persons using a similar design. Performance was lower in the elderly group, but both younger and older subjects showed the mirror effect. Experiment 3 produced confusion between preexperimental familiarity and intraexperimental familiarity by mixing traditional and novel tunes in the study lists and tests. Here, the subjects in both age groups resembled the patients of Experiment 1 in failing to show the mirror effect. Older subjects again performed more poorly, and they differed qualitatively from younger subjects in setting stricter criteria for more nameable tunes. Distinguishing different sources of global familiarity is a factor in tune recognition, and the data suggest that this type of source monitoring is impaired in AD and involves different strategies in younger and older adults.

Investigation of the cognitive deficits present in both normal and pathological aging has been occupying the energies of cognitive psychologists for a number of years. One question relating these two areas is whether the memory deficits observed in Alzheimer's disease (AD) are equivalent to those one would find in an accelerated aging process (a continuum view), or whether brain changes characteristic of AD result in a pattern of memory performance qualitatively different from that of healthy adults, both young and old. Physiological evidence consistent with the continuum view includes the fact that the neurofibrillary tangles and neuritic plaques present in the cortex of AD patients, which help define the illness pathologically (Dama-

sio, Van Hoesen, & Hyman, 1990), are present to some degree in many elderly brains (Duara, London, & Rapaport, 1985). Psychological evidence consistent with the continuum view includes the fact that tasks involving effortful retrieval of newly learned information are subject to deficits in normal elderly populations and are often the first tasks to show a deficiency in early stage AD, particularly with retention intervals longer than a few minutes (see Becker & Lopez, 1992, on AD; Light, 1991, on normal aging).

Despite this evidence for a continuum of deficits, other research suggests that AD results in a different type of memory disorder than that seen in normal aging. Once past the earliest stages of the disease, many AD patients seem to exhibit problems in tasks requiring access to well-learned information in semantic memory, such as listing items from a category, confrontation naming, or semantic priming (e.g., Cronin-Golomb, Keane, Kokodis, Corkin, & Growdon, 1992; Salmon, Shimamura, Butters, & Smith, 1988). Healthy elderly people, in contrast, seem to manage significantly better on such semantic memory tasks than on tasks of retrieving newly learned information (i.e., episodic memory tasks, Light, 1991). In fact, they exceed young adults on some semantic memory tasks, such as defining words on vocabulary tests. This pattern weighs against the continuum view, suggesting that elderly persons (relative to young adults) suffer one type of deficit,

This research was supported by Grant 1-R01-AGO9965 from the National Institute on Aging. We would like to thank Scott Lipscomb for his expertise in composing the derivations and for recording the stimulus tapes. He also assisted in testing subjects and analyzing data along with Melinda Andrews, Glenn Gesek, and Julene Johnson. We are grateful to Jeffrey Cummings, Pat Kesslak, and Michael Mahler for referring patients to us, and to Jill Shapira for help in contacting families. Kathleen Terlau provided an opportunity to test pilot subjects at OPICA Adult Daycare Center. Finally, we thank the patients and their families for participating in this research. Correspondence should be addressed to J. C. Bartlett, School of Human Development, GR4.1, University of Texas at Dallas, Box 830688, Richardson, TX 75083-0688 (e-mail: jbartlet@utdallas.edu).

whereas AD patients (relative to age-matched controls) suffer another type of deficit or suffer two deficits instead of just one. However, one might sustain a continuum view with evidence that a single type of deficit in retrieving well-learned knowledge creates problems: (1) primarily in episodic memory tasks (when the deficit is mild, as in healthy old age), or (2) in both episodic and semantic memory tasks (when it is more severe, as in AD). Indeed, Lars Bäckman and his associates have supported this view in a series of studies of recognition memory for faces (summarized in Herlitz, Lipinska, & Bäckman, 1992).

One relevant study was reported by Bäckman (1991), who showed young and elderly subjects photographs of people who were either famous recently (more familiar to young people) or from the 1930s and 1940s (more familiar to older people). In addition, he used unfamiliar faces that were either old or young, assuming that people would have more general knowledge about same-age versus different-age faces (see also Bartlett & Leslie, 1986; Fulton & Bartlett, 1991). The subsequent test was that of old/new recognition; subjects viewed a series of faces, judging each face as viewed previously ("old") or not ("new").

With famous faces, Bäckman (1991) found that younger subjects were indeed superior in recognizing contemporary versus dated items, and vice versa for the older subjects. For unfamiliar faces, the young and moderately old subjects (60–69 years old) were also superior on same-age faces, but the older subjects (over 70 years old) recognized old and young faces equally well (and more poorly overall than did the younger groups). Bäckman concluded that the oldest subjects could use preexisting knowledge to aid memory, but only when such knowledge was "rich" (i.e., when it included highly specific information about famous faces). With unfamiliar faces, the only knowledge that is relevant consists of general information about the appearance of faces in various classes (e.g., the faces in one's age group). The oldest adults seemed unable to use this more impoverished information to assist them in the episodic task.

In a related study with AD patients, Bäckman and Herlitz (1990) gave dated and contemporary famous faces to normal elderly people and AD patients for recognition 15 min later. Both groups showed greater knowledge of the dated versus contemporary faces, as indicated by a greater number of "familiar" judgments and correct name recognitions for the dated faces. However, in old/new recognition, only the unimpaired people showed the expected superior performance for the dated faces. Patients performed more poorly than did controls, and they showed equal memory performance with the two types of faces. The authors concluded that one symptom of AD is an inability to use task-relevant prior knowledge to aid episodic remembering, even when that knowledge is relatively rich. However, in a follow-up study, Lipinska, Bäckman, and Herlitz (1992) used procedures that encouraged more elaborative encoding at the time of study (faces were accompanied by their names, and subjects attempted to generate facts about each person). In that study, the AD patients showed improved performance, as well as the theoretically critical advantage of dated faces over con-

temporary faces. This outcome suggested that AD patients are not completely unable to use their prior knowledge to aid episodic memory; they simply require more direction at the time of study than do control subjects.

One view of the work by Bäckman and his colleagues is that it demonstrates a continuum of memory deficits; moving from youth to early old age to later old age to Alzheimer's disease, there is a decreasing ability (or spontaneous inclination) to utilize preexisting knowledge to aid encoding and retrieval in episodic memory tasks. Even very elderly people use such knowledge spontaneously so long as it is rich, though not if it is relatively impoverished. Early stage AD patients need substantial direction at encoding to use even relatively rich representations of knowledge. This work also suggests that knowledge about prominent individuals, at least in sports, politics, and entertainment, stays fairly intact throughout a normal lifespan and also in the early stages of AD. The deficits pertain to the use of knowledge rather than to its presence or absence.

While these data tend to support a continuum view, an intriguing result raises questions about it. In the articles just cited, Bäckman and colleagues reported inferential statistics only on the measure of d' . However, a close examination of average hits and false alarms (which were reported in the articles) reveals some informative trends. For example, in Bäckman and Herlitz's (1990) study comparing AD patients with age-matched controls, the latter group showed an increase in hit rate from .87 for contemporary faces to .94 for dated faces, along with a drop in false-alarm rate from .04 to .01. This pattern, known as the *mirror effect* (Glanzer & Adams, 1985), is often observed in recognition memory for more and less memorable stimuli. However, the AD patients in the Bäckman and Herlitz study showed what has been described as the "opposite" of the mirror effect (see Glanzer & Adams, 1985, p. 8): While the patients' hit rate increased (from .66 for contemporary faces to .74 for dated faces), so did their false-alarm rate (from .32 to .37).

Bäckman and Herlitz's (1990) results suggest that though recognition memory in AD patients shows departures from the mirror effect, recognition memory in healthy elderly persons does not. Indeed, this suggestion was buttressed in Bäckman's (1991) study, in which all three groups of healthy elderly participants in each of two experiments showed higher hit rates and lower false-alarm rates for dated items (faces, names, and face-name pairs) than for contemporary items. The picture is clouded to some extent in Lipinska et al. (1992), where the mirror effect was not supported for AD patients or for (elderly) controls. Notwithstanding this last study, the bulk of Bäckman's results suggests that the mirror effect with more (vs. less) familiar stimuli is supported just as well for healthy older persons as for young adults, but that, in some cases, it breaks down in AD. This tentative conclusion would appear to pose problems for a continuum view of normal and pathological aging, since it suggests there may be a qualitative difference between AD patients and healthy young and older persons in how recognition judgments are made.

Were such a qualitative difference confirmed, what would it imply? The pattern of performance shown by the AD subjects in the Bäckman and Herlitz (1990) study conforms to well-known memory models that “cast recognition memory in signal-detection terms, as a discrimination that is based on a single criterion on a unidimensional strength or familiarity scale” (Hintzman, Caulton, & Curran, 1994, p. 275). As illustrated in Figure 1, such unidimensional familiarity models imply that familiarity based on preexperimental knowledge and intraexperimental “oldness” should tend to have parallel (noninteracting) effects, so that hit rates as well as false-alarm rates should be higher for more familiar stimuli than for less familiar stimuli. This was the pattern shown by the AD subjects in the Bäckman and Herlitz study.

Unlike the performance of the AD subjects, that of the healthy young and older persons in Bäckman (1991) and Bäckman and Herlitz (1990) does not fit the unidimensional familiarity model. These data showed the mirror effect of increased hits and reduced false alarms for more familiar stimuli than for less familiar stimuli, suggesting two general classes of account, which are illustrated in Figures 2 and 3. First, healthy subjects might base old/new judgments on a variable other than global familiarity, a variable on which high-familiar items are relatively low when they are “new” and relatively high when they are “old” (Figure 2). The nature of this variable and the nature of the processes through which it is derived are currently unknown and under discussion (see, e.g., Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Hintzman, 1994).

Second, while healthy subjects might base old/new judgments on global familiarity, this would occur only following some adjustment of recognition criterion, or familiarity itself (cf. Brown, Lewis, & Monk, 1977; Hintzman et al., 1994).

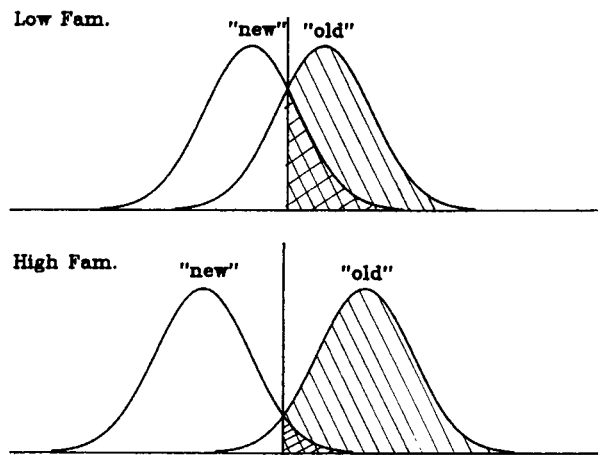


Figure 2. Distributions of a variable signifying “oldness” for “old” and “new” items whose preexperimental familiarity is low (upper distributions) or high (lower distributions). Responses based on the symmetrically placed criterion will exhibit the mirror effect of higher hit rates and lower false-alarm rates for the more familiar items. Note that the more familiar stimuli are higher on the “oldness” variable when they are “old,” and yet they are lower on this variable when they are “new.” Hence, the “oldness” variable in this case is not global familiarity.

man et al., 1994). As shown in Figure 3, global familiarity must certainly be higher for preexperimentally well-known stimuli than for preexperimentally poorly known stimuli. However, judicious shifts in the recognition criterion, or in the global familiarity of some item types, could produce the finding of lower false-alarm rates as well as higher hit rates for preexperimentally well-known items than for preexperimentally poorly known items.

Of course, if subjects use an adjustment process, they would require some way of deciding if an adjustment to criterion or familiarity were needed and, if so, in what direction as well as how much. For example, being able to identify a face verbally is evidence that the face is well known in life and that its perceived familiarity can be attributed in part to preexperimental knowledge. Hence, a subject who is taking an old/new recognition test might use a higher criterion when a face can be labeled than when it cannot (see panel A of Figure 3). If we add the assumption that studying a face increments its global familiarity, and that this effect is greater for better known faces (due to more elaborative encoding), the mirror effect could be explained.

Alternatively, a failure to retrieve episodic information concerning prior study of a face might be the basis for an adjustment. Hence, when a face that is high in global familiarity does not evoke retrieval of episodic information, the subject might attribute its high familiarity to experiences in life instead of to the input list. A downward adjustment in the item’s familiarity might be called for in this case (see panel B of Figure 3). In contrast, when a face that is moderate in global familiarity does not evoke retrieval, it might be judged as “old” but simply poorly encoded. In that case, no downward adjustment would be

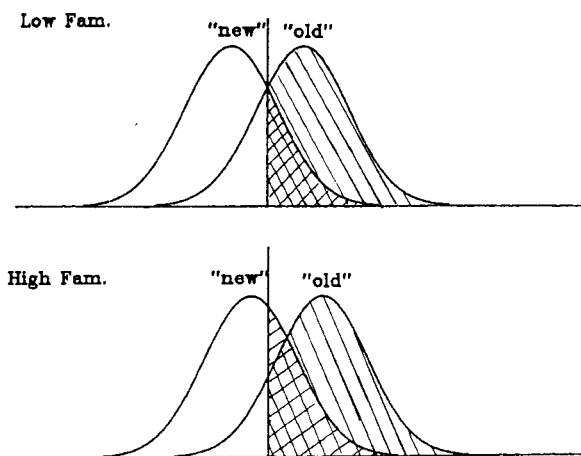


Figure 1. Distributions of global familiarity for “old” and “new” items whose preexperimental familiarity is low (upper distributions) or high (lower distributions), according to a unidimensional familiarity model. Note that the more familiar stimuli are higher in global familiarity and that the location of criterion is fixed. Hence, both hit rates and false-alarm rates should be higher for the more familiar items.

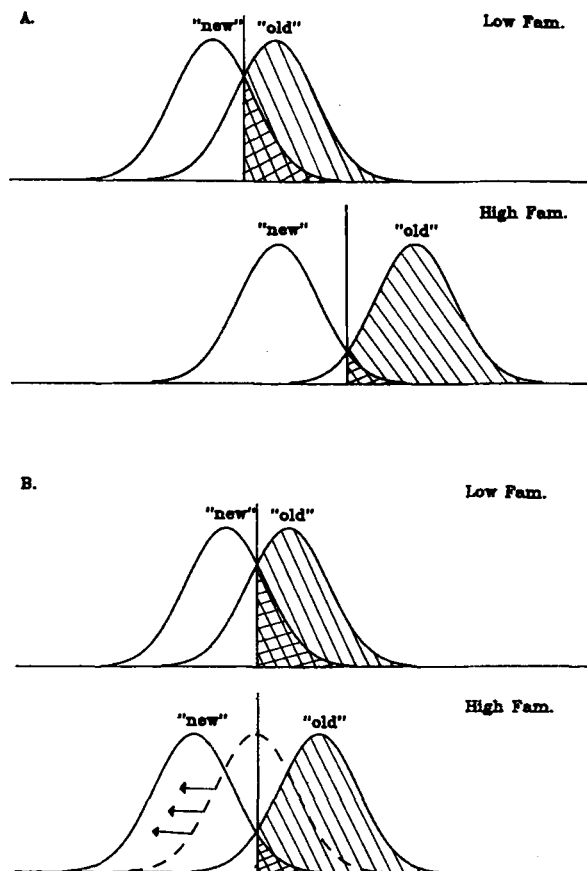


Figure 3. Panel A shows distributions of global familiarity for “old” and “new” items whose preexperimental familiarity is low (upper distributions) or high (lower distributions), illustrating the hypothesis that old/new judgments are based on global familiarity and that subjects shift their criteria for more and less familiar items. Note that though familiarity is higher for high-familiar items than for low-familiar items, responses based on the differing criteria will exhibit the mirror effect of higher hit rates and lower false-alarm rates for the former than for the latter. Panel B shows similar pairs of distributions, illustrating the hypothesis that old/new judgments are based on global familiarity and that subjects achieve the mirror effect through making downward adjustments to global familiarity for high-familiar “new” items based on failures to retrieve episodic information.

made. Again, the mirror effect of increased hits and reduced false alarms for more familiar items than for less familiar items could result from this process.

There may be other strategies besides the two just mentioned for adjusting criteria or global familiarity to produce the mirror pattern of more familiar stimuli evoking lower false-alarm rates as well as higher hit rates. However, any adjustment strategy that can produce the mirror pattern must be based on the ability to make valid attributions about the source of one’s feelings of global familiarity. Without reasonable accuracy in this type of “source monitoring” (Jacoby, Kelley, & Dywan, 1989), subjects who are using global familiarity could not know how to

shift their criteria or alter the level of familiarity itself to achieve the mirror pattern.

In summary, the failure of the mirror effect that can occur in AD is consistent with the use of global familiarity and an absolutely fixed criterion in old/new recognition. The occurrence of the mirror effect in healthy young and older subjects could arise in two ways: (1) through use of a variable other than global familiarity, or (2) through adjustments to criterion or to familiarity itself based on source monitoring. In either case, AD patients would suffer from a deficit different in kind from that linked to normal aging.

To examine further whether AD-related deficits and age-related deficits fall on a continuum, the present Experiment 1 examined the mirror effect among AD patients and age-matched controls using a stronger manipulation of preexperimental familiarity than that employed by Bäckman and Herlitz (1990). In addition, Experiments 2 and 3 looked at the mirror effect among healthy young-adult and elderly subjects under conditions designed to produce different frequencies of source-monitoring errors. If the mirror effect with more and less familiar stimuli depends on adjustments to criterion or global familiarity based on source monitoring, departures from the mirror effect should be found in conditions that cause errors in that source monitoring.

An additional goal pursued in this research was to generalize the work in recognition memory for high- and low-familiar materials to melodic stimuli. Anecdotal evidence is reported from time to time that some AD patients retain musical skills, memory, or appreciation after other abilities have deteriorated (Beatty, Zavadil, & Bailly, 1988; Crystal, Gober, & Masur, 1989). A few studies in the music therapy literature (Clair & Bernstein, 1990; Norberg, Melin, & Asplund, 1986; Olderog-Miller & Smith, 1989; Smith, 1986) claim that musical activities elicit positive responses even from severely impaired people. However, aside from these clinical reports and speculations, no systematic exploration of music cognition in AD or even normal aging has so far been attempted (a study of music perception in healthy elderly by Halpern, Bartlett, & Dowling, 1995, is an exception). If musical skills are preserved in the early stages of dementia and/or in normal aging, we would want to consider what factors might make music a special case. On the other hand, if musical abilities (like other abilities) are adversely affected in dementia and/or aging, we will have additional evidence of the global nature of certain cognitive deficits.

Our paradigm was based on the work of Bäckman and Herlitz (1990). Whereas they used dated and contemporary faces to vary familiarity among their elderly participants, we were able to vary familiarity more extremely by using common “traditional” tunes versus newly composed “novel” tunes as our stimulus materials. As described below, we were careful to make the novel tunes musically equivalent to the traditional tunes, which were familiar holiday, children’s, and patriotic melodies. One other difference from the earlier work was that, in some experiments, we presented tunes blocked by familiarity condi-

tion rather than with familiar and unfamiliar items mixed within the study list (as did Bäckman and Herlitz). Although we did this initially in Experiment 1 to make the task easier for the AD patients, Experiments 2 and 3 contrasted blocked and mixed presentation to manipulate the difficulty of source monitoring.

EXPERIMENT 1

Normal and demented subjects first heard a series of traditional tunes, followed by an old/new recognition test. They then heard a list of novel tunes and again took an old/new recognition test. Finally, the subjects heard the entire set of melodies played in a random order. They judged each tune as either familiar or unfamiliar and attempted to recall the title, lyrics, or tune category for each tune judged to be familiar.

Our interest was focused on three predictions: First, if AD patients retain their knowledge about traditional tunes, they should perform fairly well, both in absolute terms and relative to control subjects, in the familiarity-rating/title-recall test. Second, if AD patients are impaired in utilizing such knowledge to aid old/new recognition (Bäckman & Herlitz, 1990), they should do relatively worse, relative to controls, in old/new recognition of traditional tunes (which they know) than in old/new recognition of novel tunes (which they do not know). A third prediction follows if AD patients' deficits in old/new recognition are due not simply to failures in using prior knowledge but to problems in attributions regarding different sources of global familiarity. In that case, AD patients and controls should show a similar pattern of hits, but only controls should make fewer false alarms with traditional tunes than with novel tunes.

Although all three predictions were supported by Bäckman and Herlitz (1990), whether these predictions would continue to be supported here was not clear for three reasons. First, generalization from faces to melodic stimuli could hardly be assured. Second, as already noted, our familiarity manipulation (a comparison of traditional versus novel tunes) was stronger than that of Bäckman and Herlitz (a comparison of more and less familiar faces). Hence, the present experiment provided a stronger test of AD-related deficits in utilizing prior knowledge. Third, whereas Bäckman and Herlitz (1990) intermingled more and less familiar faces within their study lists and tests, we presented and tested traditional and novel melodies blocked by tune type; a study list and test of traditional tunes was followed by a study list and test of novel tunes. Although, initially, we planned to present traditional and novel tunes mixed, pilot testing showed that this procedure was unworkable for our impaired participants. Despite careful instructions to the contrary, several AD pilot subjects made responses of "old" for all traditional tunes, and "new" for all novel tunes, whether presented earlier or not. This outcome in itself was encouraging evidence that AD patients could use preexperimental knowledge to distinguish familiar tunes from unfamiliar tunes and that they base old/new judgments on global familiarity. However,

an alternative view is that our AD patients simply misunderstood the task, failing to comprehend that only tunes heard previously in the study should be judged "old." The blocked design of Experiment 1 was intended to ensure comprehension of instructions.

Method

Subjects. The AD patients for this experiment were 15 individuals (10 men and 5 women) diagnosed as having probable Alzheimer's disease by the medical staff at the UCLA Medical Center, the West Los Angeles Veterans Administration Hospital, or the UC-Irvine Medical Center. Medical workups to exclude other possible causes of dementia included a complete neurological examination, a medical history, laboratory tests, a neuropsychological battery, and an MRI.

Patients having a severe hearing impairment or psychiatric symptoms or patients taking drugs judged to interfere with cognitive function were excluded. All patients met the NINCDS-ADRDA criteria for probable Alzheimer's disease (McKhann et al., 1984). The patients were considered to have mild to moderate levels of dementia; their Folstein Mini Mental State Exam (MMSE) scores ranged from 15 to 25, with a mean of 19.9 ($SD = 3.0$). Ages ranged from 63 to 90 years, although only 2 patients were older than 79 years ($M = 73.6$, $SD = 7.2$). Consent to participate in the study was obtained both from the patients and from an accompanying family member.

Control subjects were healthy individuals (5 men and 9 women) drawn from patients' spouses and other caregivers. The mean age of this group (range = 59–80 years, $M = 72.0$, $SD = 7.3$) did not differ significantly from that of the AD group. Likewise, years of formal education did not differ between the groups ($M_s = 14.9$ for controls, 14.5 for patients). All participants were native speakers of English brought up in the United States. A few people in each group had some musical training, but none were professional musicians.

Materials. The traditional tunes were the opening phrases of 16 melodies familiar to people of all ages in the United States. Examples included Christmas tunes ("Deck the Halls"), children's tunes ("London Bridge is Falling Down"), folk tunes ("On Top of Old Smokey"), and patriotic tunes ("America the Beautiful"). These songs have been rated as highly familiar in previous research (Halpern, 1984) and in pilot work for the present project.

The design required that the novel tunes have the same musical characteristics as the traditional tunes. Objectively, they needed to have the same number of notes as the traditional tunes and pitch intervals and rhythmic units similar to those of the traditional tunes. Subjectively, they needed to have the simple, pleasant character of the traditional tunes. This was accomplished by asking a professional musician to compose a derivation of each traditional tune. For each tune, he permuted its rhythmic units and pitch intervals to form a pleasant melody whose origin was nevertheless unrecognizable. Each derivation was played at the same tempo as the parent tune. To test the recognizability of the derivations, UCLA students listened to each derivation and noted whether it reminded them of any real tunes. Adjustments were made to derivations until we were confident that the 16 derivations, while pleasant and musically equivalent to the traditional tunes, did not resemble their parent tunes. An example of a traditional tune and the novel tune made from it is shown in Figure 4.

All tunes were synthesized on a Kawai KLM module using the Piano 1 setting and were sequenced on a Yamaha QX1 sequencer. Tunes were recorded on audio tape and presented to participants via a high-quality stereo cassette player. Each excerpt was approximately 5 sec long. In the presentation phase, the interstimulus interval was 4 sec. In the test phase, the response interval was 7 sec.

The experimental sequence consisted of 8 traditional tunes, followed by a recognition test of 16 traditional tunes (8 old and 8 new). This was followed by 8 novel tunes and the corresponding test of 16 novel tunes. This blocked arrangement was always presented in the order traditional–novel. We felt that the reverse order might be con-



Figure 4. An example of a traditional tune and a novel tune that was derived from it.

fusing and frustrating for the patients. For counterbalancing purposes, two tapes were prepared such that an item serving as a recognition target in Tape A served as a foil in Tape B. Half of the subjects in each group heard each tape.

Procedure. Participants were tested individually in a quiet room, usually as part of a routine clinic visit for the patient. Instructions stated:

I am going to play some tunes for you. They will be familiar to you (you will recognize them). You don't have to sing or name them. Listen to them carefully and try to remember them. I will ask you questions about them later.

The 8 traditional tunes were then played without interruption followed immediately by the recognition test. Instructions stated:

Now I'm going to play some more tunes. Some of these you just heard now, but some of them were not on the tape you just heard. As I play each tune, please say "yes" if you just heard the tune on the tape, and "no" if you did not just hear it. Remember, even though all of these are familiar tunes, some you just heard on the tape and others you didn't, so sometimes the correct answer is "yes" and sometimes the correct answer is "no."

The average retention interval for tunes (i.e., the time between presentation of a tune and its appearance on the recognition test, including the time for reading the instructions) was approximately 3 min (range = about 1–4 min). The participants stated their answers aloud; the experimenter prompted for answers if necessary. The experimenter also paused the tape occasionally when more time was needed.

After the recognition test with the traditional tunes, the experimenter gathered some demographic data, such as age, years of education, and musical background. The unfamiliar tunes were then presented, using the same timing as that described above, and using similar instructions with the appropriate wording changes.

After the subjects completed the second (novel-tunes) test, they performed a familiarity-rating and title-recall task. All 32 tunes used in the experiment were presented one at a time, in a mixed list. Instructions were to listen to each tune and to indicate by a "yes" or "no" whether the tune was known before the experimental session. In addition, for tunes that were indicated as being familiar, we asked the listener to name the tune or its first few words. If someone indicated familiarity but could not retrieve a title or lyrics, he or she was given a forced choice of tune category: patriotic, special occasion or holiday, children's, or folk tune.

To monitor whether the memory test was having a deleterious effect on the patients' well-being, we asked each patient to rate his or her mood on a scale of 1 (*sad*) to 10 (*happy*) at three points in the study: after each of the recognition tests and after the familiarity ratings. These mood ratings were also available for 8 of the 14 normal controls. At the conclusion of the session, the participants were reassured that half of the tunes ought not to have been familiar because they were newly composed. Experimental sessions lasted approximately 30 min.

Results

Mood ratings. All participants seemed to enjoy the study. During the familiarity ratings, many people hummed

along or commented on the particular tunes. The patients did not seem to be distressed by their poor memory performance. On the contrary, mood ratings indicated that all subjects improved their moods throughout the session: $M_s = 6.9, 7.4, 7.7$ (out of 10) for the three times of measurement [$F(2,40) = 5.88, MS_e = 0.67, p < .01$], with no reliable group effect or group \times time interaction.

Familiarity judgments and naming of tunes. Before examining performance in the old/new recognition tests, it is useful to inspect the data on familiarity ratings and naming of the tunes collected at the end of the experimental session. The proportions of correct "yes" ("familiar") judgments made to traditional songs (hit rates) averaged .99 ($SD = .03$) and .96 ($SD = .06$) for the control and patient groups, respectively. The proportions of erroneous "yes" judgments made to novel songs (false-alarm rates) averaged .03 ($SD = .05$) and .12 ($SD = .18$) for the two groups. A small deficit in the patient group was suggested by their slightly lower hit rates and higher false-alarm rates. However, the group differences were reliable at our chosen alpha level of .05 *only* by a one-tailed test [$t(27) = 1.9$, for hit rates; $t(27) = 1.8$, for false-alarm rates].

In scoring the naming data, 1 point was given for either a title or a few words that were clearly from the song (e.g., "merrily, merrily" or "down the stream" in response to the tune "Row Row Row Your Boat"). A subject received 0.5 point for identifying the song's category (e.g., "nursery tune") or for providing words for a song in the appropriate category (e.g., "over the bounding main"). The maximum score was 16. The control subjects scored an average of 14.2 ($SD = 2.03$), whereas the patients performed significantly worse, with an average score of 9.7 ($SD = 3.52$) [$t(27) = 4.18$]. Three patients achieved scores as good as the average control, and 1 control subject scored worse than the average patient. For the patients, naming scores and MMSE scores correlated at $r = .38$ ($df = 13$, n.s.), suggesting at best a modest relationship between degree of impairment and naming ability.

Figure 5 shows the mean naming scores for each traditional song for both subject groups. The control subjects named most of the songs, and even the patients averaged at least 0.5 point for nearly all the songs, indicating some knowledge of their identity. Performance between the two groups was significantly correlated across songs [$r(14) = .57, p < .02$].

Old/new recognition. Table 1 displays proportions of correct "old" judgments in response to old tunes (hit rates)

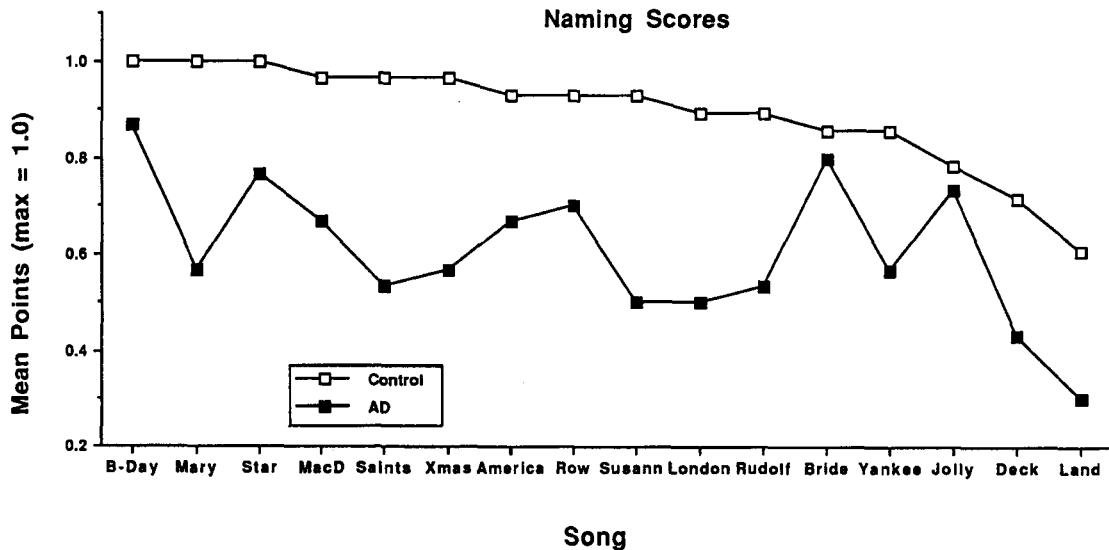


Figure 5. Naming scores for each traditional tune for the control and AD groups in Experiment 1.

and erroneous “old” judgments in response to new tunes (false-alarm rates) in the traditional-tunes test and the novel-tunes test. The control subjects showed the mirror effect, in that their hit rates appeared higher and their false-alarm rates appeared lower with traditional tunes than with novel tunes. In contrast, the patients failed to show this effect; although their hit rates appeared higher with traditional tunes than with novel tunes, their false-alarm rates did not reverse the pattern. In consequence, the hit rates suggested an AD-related deficit regardless of tune type, whereas the false-alarm rates supported an AD-related deficit only with the traditional tunes. An analysis of variance (ANOVA) of the hit rates supported only the main effect of group [$F(1,27) = 6.35, MS_e = 6.04$]. In contrast, an ANOVA of the false-alarm rates supported a main effect of tune type [$F(1,27) = 5.29, MS_e = 3.41$], qualified by a tune type \times group interaction [$F(1,27) = 6.84, MS_e = 3.41$]. The pattern resembles that found by Bäckman and Herlitz (1990).

Signal-detection-theory analysis. In assessing different patterns of hits and false alarms vis à vis the mirror effect, Hintzman et al. (1994) recommend computation of

signal-detection-theory measures of discrimination and bias, such as d' and c (see Macmillan & Creelman, 1991). Referring to the old- and new-item distributions shown in Figures 1–3, note that d' reflects the distance between a pair of distribution means, whereas c reflects the criterion's location relative to the cross-point of the pair of distributions. A c score of zero indicates that the criterion is located at the cross-point; a positive c indicates that the criterion is located to the right of this cross-point (indicating a conservative bias); and a negative c indicates that the criterion is located to the left (indicating a liberal bias).

Since the mirror effect involves a symmetrical change in hits and false alarms, it can be defined in the context of signal-detection theory as a pattern in which two conditions differ in d' but not c . Hence, the hypothetical situations shown in Figures 2 and 3 all conform to the mirror effect. Specifically, all illustrate the case in which d' s are greater for more familiar items than for less familiar items, whereas c is always zero. Conversely, Figure 1 illustrates a departure from the mirror effect in which d' is the same for more and less familiar items, whereas c becomes negative for more familiar items. (Since the distributions for more familiar items are shifted to the right, the [fixed] criterion is no longer at the cross-point and produces a negative c for these items.)

Taking the signal-detection approach, we computed the standard discrimination index (d'), as well as a bias measure (c), for each tune type and each individual subject. The d' s of the control subjects averaged 2.14 and 0.88 for traditional and novel tunes, respectively (SD s = 1.14 and 1.34), whereas the d' s of the patients averaged 0.66 and 0.41 for traditional and novel tunes, respectively (SD s = 0.67 and 0.98). These means support the advantage of the control subjects over the patients seen in hit rates, as well as the interactive pattern seen in false-alarm rates. Indeed, an ANOVA supported reliable effects of group [$F(1,27) =$

Table 1
Hit and False-Alarm Rates in Old/New Recognition of Traditional and Novel Tunes by Control Subjects and Patients in Experiment 1

Tune Type	Controls (n=14)		Patients (n=15)		Difference
	M	SD	M	SD	
Hit Rates					
Traditional	.75	.14	.59	.27	.16*
Novel	.67	.20	.50	.25	.17*
False-Alarm Rates					
Traditional	.14	.10	.37	.15	-.23†
Novel	.38	.27	.35	.27	.03

*.05 < p < .10, by t test. † p < .001, by t test.

10.9, $MS_e = 1.27$] and tune type [$F(1,27) = 8.36$, $MS_e = 0.95$], and the group \times tune type interaction just missed significance [$F(1,27) = 3.96$, $MS_e = 0.95$, $p < .06$]. Note that the advantage of traditional over novel tunes was reliable in the control group [$t(13) = 3.09$], but not in the patient group ($t < 1$).

The c scores suggested that the control subjects were more conservative with traditional tunes ($M = +.24$, $SD = .41$) than with novel tunes ($M = -.07$, $SD = .54$), whereas the patients were more liberal with traditional tunes ($M = +.03$, $SD = .54$) than with novel tunes ($M = +.32$, $SD = .87$). The pattern is that of a cross-over interaction, which was reliable by ANOVA [$F(1,27) = 4.63$, $MS_e = 0.276$]. The tune-type effect was not statistically reliable in either subject group, but it is fair to conclude that the AD patients, compared with the controls, were relatively more liberal with traditional tunes than with novel tunes.

Discussion

Familiarity ratings suggested only minimal and marginally reliable AD-related deficits in knowledge of our tunes. The title-recall data supported somewhat stronger deficits, in agreement with prior evidence that AD patients are impaired in effortful retrieval of verbal/semantic knowledge (Cronin-Golomb et al., 1992; Salmon et al., 1988). These effects notwithstanding, more robust group differences were found in old/new recognition.

Hit rates in old/new recognition supported an AD-related deficit with traditional tunes as well as novel tunes. The false-alarm rates in old/new recognition showed an interactive pattern in that the AD-related deficit was restricted to traditional tunes. As in Bäckman and Herlitz (1990), the control subjects were successful in reducing false-alarm rates for well-known items, relative to poorly known or unknown items, whereas the AD subjects were not. Indeed, the reduced false-alarm rates shown by the controls produced important differences between the subject groups in the pattern of discrimination accuracy (d' scores) as well as bias (c scores).

Regarding discrimination, the advantage of traditional tunes over novel tunes was stronger in the control group than in the AD group. Although the relevant interaction was only marginally reliable ($p < .06$), it replicates the pattern found by Bäckman and Herlitz (1990), extending their conclusion that AD patients benefit less from prior knowledge about stimuli in old/new recognition than do controls.

Regarding bias, we found that the AD patients, relative to the controls, showed relatively more liberal bias with traditional tunes than with novel tunes.¹ Bäckman and Herlitz (1990) did not report bias scores, but c scores computed from their hit and false-alarm rates showed a similar pattern: These scores suggest more liberal bias with more familiar stimuli than with less familiar stimuli among their AD subjects ($cs = -.16$ and $+.03$, respectively), but not among their control subjects ($cs = +.39$ and $+.31$, respectively).

Bäckman and Herlitz (1990) concluded from their data (primarily their d' s) that AD patients are unwilling or unable to utilize prior knowledge to aid their performance in

old/new recognition. However, a more comprehensive view of the accumulated findings is that AD patients utilize their prior knowledge at least to the extent that they perceive well-known stimuli as more globally familiar than poorly known or unknown stimuli, as shown in Figure 1. This unidimensional familiarity hypothesis correctly predicts that AD subjects should make relatively large numbers of erroneous "old" judgments to recognition-test lures when these lures are highly familiar in life. For the same reason, it correctly predicts that while AD subjects should show similar d' s for more familiar stimuli versus less familiar stimuli, they should be relatively more liberal in making "old" judgments to the former than to the latter.

EXPERIMENT 2

Experiment 1 confirmed prior evidence (Bäckman & Herlitz, 1990; Lipinska et al., 1992) that old/new recognition is impaired in AD; music recognition is not spared in this regard. It also supported a previous indication (Bäckman & Herlitz, 1990) that recognition memory in AD patients differs qualitatively from that of controls. As in the prior study, the controls, but not the patients, showed lower false-alarm rates for stimuli that were high in preexperimental familiarity than for those that were low. In consequence, the AD patients, compared with the controls, were relatively more liberal in making "old" judgments to traditional melodies than to novel melodies.

Another prior finding was that old/new recognition was impaired in healthy elderly persons, compared with young adults (Bäckman, 1991). This age-related deficit appeared more quantitative than qualitative; although old/new recognition was generally less accurate among elderly subjects than among younger subjects, all age groups showed the mirror pattern of higher hit rates and lower false-alarm rates for more familiar stimuli than for less familiar stimuli.

The goal we pursued in Experiment 2 was to extend the evidence for this quantitative age deficit, generalizing the finding from faces to music. The study was identical to Experiment 1, except that (1) the subjects were healthy elderly persons and young adults (instead of AD patients and age-matched controls), and (2) test order was counterbalanced such that about half of the subjects received the traditional-tunes test prior to the novel-tunes test (as in Experiment 1) and the remainder received the opposite test order. We expected that performance of the elderly subjects would resemble that of the controls (also healthy elderly persons) used in Experiment 1. Indeed, we hoped that our counterbalancing of tune type and test order might produce a clearer mirror pattern than what we previously observed. We also predicted that though the young adults would outperform the elderly, the mirror pattern would not vary with age.

Method

The 26 young adult subjects (20–29 years of age, $M = 23.9$, $SD = 2.7$) were undergraduate psychology students at the University of Texas at Dallas. They participated as a means of fulfilling a psychology course requirement. The 30 elderly adult subjects (61–78 years of age, $M = 69.5$, $SD = 4.5$) were recruited from church groups in the Dallas area and participated for pay (\$10 for the 1-h session).

Scores on a vocabulary test (second half of the WAIS) averaged 23.7 and 29.8 out of 40 ($SDs = 6.8$ and 6.9) for the young and elderly subjects, respectively. The age difference was reliable [$t(54) = 3.38$], suggesting higher verbal skills among our older subjects. According to a musical background questionnaire, 12 young subjects and 15 elderly subjects had 0–2 years of formal training on an instrument or voice, 9 young and 6 elderly subjects had 3–8 years of training, and 5 young and 9 elderly subjects had more than 8 years of training.

The procedures and materials were identical to those of Experiment 1, except about half of the young and elderly subjects received the traditional-tunes test before the novel-tunes test, whereas the remainder received the novel-tunes test first.

Results

Familiarity judgments and naming of tunes. As in Experiment 1, we begin by examining the tune familiarity and naming data collected at the end of the experimental session. The proportions of correct “yes” (“familiar”) judgments made to traditional songs (hit rates) were virtually at ceiling; they averaged .98 and .99 ($SDs = .03$ and $.01$) for young and elderly subjects, respectively. The proportions of erroneous “yes” judgments made to novel songs (false-alarm rates) averaged .09 and .05 ($SDs = .10$ and $.12$) for young and elderly subjects, respectively. Naming scores averaged 14.3 and 14.0 out of 16 ($SDs = 1.67$ and 2.52) for the young and elderly subjects, respectively. An ANOVA performed on the false-alarm rates and another performed on the naming scores supported no reliable ($p < .05$) effects involving age or test order (traditional–novel vs. novel–traditional).²

Old/new recognition. Table 2 displays proportions of correct “old” judgments to old tunes (hit rates) and erroneous “old” judgments to new tunes (false-alarm rates) in the traditional-tunes test and the novel-tunes test. There was strong support for a mirror effect in that hit rates were higher and false-alarm rates were lower with traditional tunes than with novel tunes. A small age-related deficit was also suggested in that hit rates appeared higher and false-alarm rates appeared lower in the young adult group than in the elderly group. However, there was no indication that the mirror effect was more strongly supported in one age group than the other or that age-related deficits varied with tune type.

The hit rates were subjected to an ANOVA that included age as a between-subjects variable and tune type (tradi-

tional vs. novel) as the within-subjects variable. The age-related deficit was supported by a main effect of age [$F(1,48) = 4.48$, $MS_e = 4.40$], and the advantage of traditional tunes over novel tunes was supported by a main effect of tune type [$F(1,48) = 30.2$, $MS_e = 3.09$]. A similar ANOVA of the false-alarm rates showed no main effect for age, but it did support the main effect of tune type [$F(1,48) = 30.5$, $MS_e = 2.04$]. We conclude that the data supported the mirror effect and also suggested a young-subject advantage in hits.

Signal-detection-theory analysis. The effects that were found in the hit and false-alarm rates were also supported by an ANOVA of d' s. There was a reliable main effect for age [$F(1,48) = 4.42$, $MS_e = 1.51$], with d' s in the young and elderly groups averaging 2.30 and 1.78, respectively. The main effect of tune type was significant as well [$F(1,48) = 49.1$, $MS_e = 1.19$], with d' s averaging 2.78 in the traditional-tunes test and 1.31 in the novel-tunes test. An ANOVA of c scores (for bias) supported no conventionally significant effects (all $ps > .10$). These scores averaged +.12 and +.17 for young and elderly subjects, respectively, and +.13 and +.17 for traditional and novel tunes, respectively. Note that the c scores did not differ reliably from 0 (i.e., no bias) with either tune type in either age group and that the effects of tune type on d' s, but not c s, fit Hintzman's definition of the mirror effect (Hintzman et al., 1994).³

Discussion

Our younger and older subjects showed equivalent knowledge of our tunes. The familiarity ratings showed near-perfect hit rates and very low false-alarm rates in both age groups, and the naming scores were not affected by age. Notwithstanding this finding, old/new recognition showed an age-related deficit; the elderly subjects were less accurate than were the young-adult subjects by the measures of hit rate and d' .

Our goal in Experiment 2 was to determine if there are age-related deficits in tune recognition and, if so, if these age-related deficits are qualitative in nature, involving any changes in the mirror pattern of increased hits and reduced false alarms for more familiar stimuli, as compared with less familiar stimuli. The findings supported an age-related deficit, but there was nothing in the way of a qualitative difference between young adults and elderly subjects; the mirror pattern was strongly supported in both of the age groups.

It is informative to compare the age-related deficits of Experiment 2 with the AD-related deficits of Experiment 1. The AD-related deficit in Experiment 1 involved a failure on the part of the patients to reduce false-alarm rates with traditional tunes. Thus, the largest group difference was in false-alarm rates for traditional tunes (Table 1). In contrast, the age-related deficit in Experiment 2 was not generally larger with traditional tunes than with novel tunes, and both young-adult and elderly subjects showed the mirror effect of lower false-alarm rates and higher hit rates with traditional tunes than with novel tunes. Indeed, the age-related differences were smallest where the AD-related

Table 2
Hit and False-Alarm Rates in Old/New Recognition of Traditional and Novel Tunes by Young-Adult and Elderly Subjects in Experiment 2

Tune Type	Young Adults ($n=26$)		Elderly Adults ($n=30$)		Difference
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
	Hit Rates				
Traditional	.87	.12	.80	.17	.07*
Novel	.70	.19	.59	.25	.11*
	False-Alarm Rates				
Traditional	.12	.14	.13	.13	-.01
Novel	.24	.20	.31	.21	-.07

*.05 < p < .10, by t test.

differences were largest—namely, in false-alarm rates with traditional tunes (Table 2). In terms of d' 's, whereas the AD-related deficit in Experiment 1 was larger with traditional tunes than with novel tunes, the (modest) age-related deficit in Experiment 2 was not qualified by tune type, and it was, if anything, larger with novel tunes (age difference = .71) than with traditional tunes (age difference = .28). Regarding c 's, the patients of Experiment 1 were relatively more liberal, as compared with the controls, with traditional tunes than with novel tunes. In contrast, the ANOVA of c scores in Experiment 2 supported no effects, and, if the means of these scores suggested anything at all, it was for the older subjects to be relatively more conservative with the traditional tunes (age difference = .10) than with the novel tunes (age difference = .02). In summary, the accumulated data suggest that while AD-related deficits are qualitative in nature and involve a departure from the mirror effect, age-related deficits reflect simply a reduction in overall accuracy and no departure from the mirror pattern.

EXPERIMENT 3

Although the findings of Experiment 2 appeared to be clearcut, we were not yet prepared to accept the implication that the mirror effect with more and less familiar stimuli invariably occurs among healthy elderly persons, or, for that matter, among healthy young adults. Our doubts can be framed in terms of Figures 2 and 3, which show two interpretations of the mirror pattern. According to the first interpretation (Figure 2), the mirror effect occurs because (healthy) adults base their old/new judgments on a variable other than global familiarity (e.g., likelihood ratios; see Glanzer et al., 1993). In contrast, the second interpretation (Figure 3) accepts the assumption that global familiarity is a primary basis for old/new recognition but holds, in addition, that subjects make adjustments to the "old" judgment criterion (panel A) or to familiarity itself (panel B) in order to achieve the mirror pattern. We previously argued that if such adjustments are to be effective, they must entail a process of attributing the source of global familiarity to preexperimental knowledge versus intraexperimental, or "episodic," information. Hence, the adjustment hypothesis has an interesting implication: To the extent that source monitoring is subject to errors, the mirror effect should be weakened or destroyed.

In fact, recent studies of the *false-fame* effect with proper names (Dywan & Jacoby, 1990) and with faces (Bartlett, Strater, & Fulton, 1991) have shown that source monitoring is subject to errors, especially among the elderly. In both of these studies, elderly persons were more likely to judge that a name or face of an unknown person actually was that of a famous person if the name or face had been previously presented than if it had not. Young-adult subjects did not show this effect, suggesting that confusions between intraexperimental sources of familiarity and preexperimental sources of familiarity are more marked among the elderly than among young adults. Notwithstanding this age

difference, it is important to note that even young-adult subjects will show false-fame effects under some test conditions (e.g., when they are distracted at test; see Jacoby et al., 1989). Thus, the appropriate conclusion to draw from these data is that though source monitoring can fail in young adulthood, it fails more readily in old age. The implication is clear: To the extent that the mirror effect depends on source monitoring, we should observe breakdowns in the mirror effect when source monitoring is made difficult, perhaps especially in old age.

To test this line of thinking, in Experiment 3, we switched from the blocked-list procedure of Experiments 1 and 2 to a mixed-list procedure. With the blocked-list procedure of the previous experiments, traditional and novel tunes were studied and tested in separate phases of the experiment. With the mixed-list procedure of Experiment 3, traditional and novel tunes were intermixed together within each study list and test; subjects heard a study list with traditional and novel tunes mixed together and then took a test containing old and new tunes of each type. This was followed by a second study list and test that also had a mixture of traditional and novel tunes.

With the blocked-list procedure (which was used in Experiments 1 and 2), the process of source monitoring should be relatively easy: One need only keep in mind that all of the items in a given test list are well known in life or not and that well-known items should have relatively high levels of global familiarity, even if they are new. In contrast, with the mixed-list procedure of Experiment 3, source monitoring must be done on a tune-by-tune basis at test. Only after hearing a tune could a subject classify it as traditional or novel, and it is likely that errors would occur in this process. For example, the subject might attempt to label the tune on the assumption that only traditional tunes would evoke recall of labels. However, such a labeling strategy would lead to misclassifications of well-known tunes whose labels are not retrieved.

If the mixed-list conditions of Experiment 3 produce errors in source monitoring, at least among the elderly subjects, we should see a breakdown in the mirror effect. The situation might resemble that shown in Figure 1, which was drawn to represent recognition in AD. The point of the figure is that if familiarity is higher for traditional tunes than for novel tunes, and if source-monitoring failures prevent compensatory adjustments to recognition criteria or perceived familiarity, then "old" judgments should be generally more frequent for traditional tunes than for novel tunes. In consequence, c scores should be lower for traditional tunes than for novel tunes, as was shown by AD subjects in Experiment 1 (recall that since c scores reflect the location of criterion relative to the cross-point of the two distributions, a rightward shift in both distributions will result in a reduction in c). Of course, we also expected that d' should be higher for traditional tunes than for novel tunes, as was found in both age groups in Experiment 2. Figure 1 could be changed to reflect this expectation by moving the old-item distribution for the more familiar stimuli somewhat farther to the right.

Method

The 21 young-adult subjects (17–29 years of age, $M = 21.8$, $SD = 3.5$) were undergraduates at Bucknell University. They participated as a means of fulfilling a course requirement. The 21 elderly subjects (61–81 years of age, $M = 70.4$, $SD = 5.1$) were recruited as in Experiment 2. Vocabulary scores averaged 23.4 out of 40 ($SD = 6.8$) in the young-adult group and 29.1 out of 40 ($SD = 7.4$) in the elderly group, showing reliably higher verbal skills among the older subjects [$t(40) = 2.60$]. The musical background questionnaire indicated that 11 of the young adults and 10 of the elderly adults had 0–2 years of musical experience, 2 of the young adults and 3 of the elderly adults had 3–8 years of experience, and 8 of the young adults and 8 of the elderly adults had more than 8 years of experience. The method was similar to that used in Experiment 2: Each of two study lists containing 8 tunes was followed by a test containing 8 “old” and 8 “new” tunes. Unlike Experiment 2, however, each study list and test contained equal numbers of traditional and novel melodies, intermixed randomly. The subjects were informed of this fact and were carefully instructed to make old/new judgments independently of whether the tunes were traditional or novel.

Results

Familiarity judgments and naming of tunes. As in Experiment 2, the proportions of correct “yes” (“familiar”) judgments made to the traditional tunes (hit rates) were at ceiling, averaging .97 ($SD = .03$) in the young-adult group and .99 ($SD = .01$) in the elderly group. The proportions of erroneous “yes” judgments made to the novel songs (false-alarm rates) were slightly above the floor, averaging .06 ($SD = .07$) and .04 ($SD = .05$) in the young adult and elderly groups, respectively. Naming scores (out of 16) averaged 14.6 ($SD = 1.7$) and 14.5 ($SD = 1.8$) for young and elderly subjects, respectively. An ANOVA performed on the false-alarm rates and another performed on the naming scores did not support the small age differences.

Old/new recognition. Table 3 displays proportions of correct “old” judgments in response to old tunes (hit rates), along with proportions of erroneous “old” judgments in response to new tunes (false-alarm rates). An age-related deficit is suggested by the fact that hit rates were higher and false-alarm rates were lower in the young-adult group than in the elderly group. A tune-type effect is shown by a hit-rate advantage for traditional tunes over novel tunes. However, false-alarm rates clearly were not lower for traditional tunes than for novel tunes, as they were in Experiment 2. Thus, the mirror effect found in Experiment 2 was not found in either age group here.

Table 3
Hit and False-Alarm Rates in Old/New Recognition of Traditional and Novel Tunes by Young-Adult and Elderly Subjects in Experiment 3

Tune Type	Young Adults ($n=21$)		Elderly Adults ($n=21$)		Difference
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Hit Rates					
Traditional	.91	.08	.82	.19	.09†
Novel	.57	.25	.41	.21	.16†
False-Alarm Rates					
Traditional	.14	.21	.28	.25	-.14*
Novel	.11	.10	.24	.22	-.13†

*.05 < p < .10, by t test. † p < .05, by t test.

An ANOVA of hit rates supported the age difference [$F(1,38) = 10.0$, $MS_e = 6.18$], as well as the advantage of traditional tunes over novel tunes [$F(1,38) = 79.3$, $MS_e = 7.60$]. There was also an interaction between test (first vs. second) and age [$F(1,38) = 5.10$, $MS_e = 2.71$]: The age difference was larger in the first test (mean difference = .18) than in the second test (mean difference = .07).

An ANOVA of false-alarm rates supported the age difference [$F(1,38) = 6.59$, $MS_e = 11.33$], but not the effect of tune type. However, there was a tune type \times test interaction [$F(1,38) = 8.87$, $MS_e = 2.67$]: False-alarm rates in the first recognition test appeared slightly lower for traditional tunes ($M = .14$) than for novel tunes ($M = .18$), though the difference was not reliable. In contrast, false-alarm rates in the second test were reliably higher for traditional tunes ($M = .28$) than for novel tunes ($M = .17$) [$F(1,38) = 4.86$, $MS_e = 5.14$], a clear violation of the mirror effect. A conservative conclusion is that the data violate the mirror pattern of a significant reduction in false-alarm errors for traditional tunes, relative to those for novel tunes, especially in Test 2.

Signal-detection-theory analysis. The d' scores for old/new discrimination were higher in the young-adult group ($M = 2.73$) than in the elderly group ($M = 1.56$), and they also were higher for traditional tunes ($M = 2.97$) than for novel tunes ($M = 1.31$). An ANOVA supported both main effects [$F(1,38) = 15.4$, $MS_e = 3.70$, and $F(1,38) = 57.5$, $MS_e = 2.02$, respectively], as well as a tune type \times test interaction [$F(1,38) = 11.6$, $MS_e = 1.44$]. The advantage of traditional tunes over novel tunes was greater in the first test (difference = 2.30) than it was in the second (difference = 1.03).

The c scores for bias in old/new recognition were substantially more liberal for traditional tunes ($M = -.13$) than for novel tunes ($M = +.73$), reflecting a departure from the mirror effect. There was a much weaker tendency for more conservative criteria in the first test ($M = +.40$) than in the second ($M = +.20$). An ANOVA supported both main effects [$F(1,38) = 44.2$, $MS_e = 0.70$, for tune type; $F(1,38) = 8.28$, $MS_e = 0.22$, for test], with no reliable interactions.⁴

Discussion

Again, neither the familiarity ratings nor the title-recall scores suggested age-related deficits in preexperimental knowledge of our tunes. However, old/new recognition showed age differences that were somewhat more marked than in Experiment 2. Mean d' scores in Experiment 3 averaged 2.73 and 1.56 for young-adult and elderly subjects, respectively. The comparable d' s in Experiment 2 averaged 2.30 and 1.78, respectively.

By far, the greatest difference between Experiment 3 and Experiment 2 pertained to the question of mirror effects. In Experiment 2, the mirror effect was supported in that both young and older subjects showed higher hit rates and lower false-alarm rates with traditional tunes than with novel tunes. Moreover, d' s were higher for traditional tunes than for novel tunes, whereas c scores showed virtually no effect of tune type, which fits the definition of the mirror

effect suggested by Hintzman et al. (1994). Turning to Experiment 3, both young and older subjects (again) showed higher hit rates with traditional tunes than with novel tunes. However, neither young nor older subjects replicated the finding of lower false-alarm rates with traditional tunes. In fact, a false-alarm difference in the opposite direction was reliable in the second test. Another difference from Experiment 2 was that though d' s were higher for traditional tunes than for novel tunes, c scores were significantly lower (i.e., bias was relatively more liberal) with traditional tunes than with novel tunes. We conclude that though the data of Experiment 2 conformed very well to the mirror effect, the data of Experiment 3 did not.

Although Experiments 2 and 3 produced differing outcomes, an adjustment hypothesis can be applied to both. If we assume that the subjects' old/new judgments were based at least partly on global familiarity, the mirror effect would depend upon a strategy, such as raising criteria and/or decrementing familiarity for some traditional tunes (or vice versa for some novel tunes, see Figure 3). Such a strategy would be easy to follow when traditional and novel tunes are studied and tested in separate lists (as in Experiment 2), but not when they are mixed together in each study list and test (as in Experiment 3). To the extent that such mixed-list conditions prevent appropriate adjustments in criterion or familiarity, the mirror effect should fail to occur. Specifically, although hit rates might be higher for traditional tunes than for novel tunes, false-alarm rates should not be lower for the former than for the latter. In consequence, although d' scores should be higher for traditional tunes than for novel tunes, c scores should not be constant across the two tune types. Instead, c scores should be reliably lower for traditional tunes than for novel tunes, as was found in Experiment 3 (and also suggested in the AD group in Experiment 1).

A question raised by the preceding discussion is that of whether the subjects in Experiment 3 used an adjustment strategy with less than complete success, or if they simply based their "old" judgments on "raw" familiarity and a fixed criterion level as shown in Figure 1. The former possibility is supported by the fact that though the findings departed from the mirror effect, the false-alarm rates were only slightly higher for traditional tunes than for novel tunes (Table 3). This was in contrast to the hit rates, which were considerably higher for traditional tunes. Considering also the fact that there was a weak hint of the mirror effect in the first recognition test, we can conclude that the data neither supported the mirror effect nor fit predictions of a unidimensional familiarity hypothesis (i.e., hit rates and false-alarm rates both higher for traditional tunes than for novel tunes; see Figure 1). Instead, the data fell between these two patterns, which suggests that the subjects used an adjustment strategy, albeit in an imperfect way.

Age differences in strategies for source discrimination. That the data fit an adjustment hypothesis is important for interpreting the age-related differences in recognition accuracy (d') found here as well as in Bäckman's (1991) research. Previously, we argued that the adjustment

hypothesis implies a process of source monitoring: if adjustments to criterion or familiarity are to produce the mirror effect, or any approximation to it, a subject's feelings of global familiarity must be attributed with some accuracy to experiences in life (preexperimental sources) as opposed to experiences in the study phase (intraexperimental episodic sources).

In light of this argument, it is important to note that the mirror effect was supported (or violated) to the same degree in the young-adult groups as in the elderly groups. From the standpoint of the adjustment hypothesis, this finding implies that, despite age differences in recognition accuracy, source monitoring is largely age-invariant under the conditions used here. However, there remains the possibility of important age differences in the strategies that the subjects used for source monitoring.

We previously considered two possible source-monitoring strategies that the subjects may have used, or attempted to use, in the mixed-list conditions of Experiment 3. A *recollection* strategy would involve an attempt to retrieve an episodic memory from the study phase in response to each test tune: If the tune is high in global familiarity and yet fails to evoke an appropriate episodic memory, familiarity would be adjusted down (Figure 3, panel B). A *labeling* strategy would involve an attempt to label or identify each test tune: Upon encountering the tune, the criterion would be adjusted up (or familiarity would be adjusted down) if the tune can be labeled, and/or vice versa if it cannot (Figure 3, panel A). The labeling strategy has an interesting limitation: It might work in some cases, but it would lead to a relatively liberal bias with traditional tunes that are difficult to name. Specifically, we should see higher rates of "old" judgments for traditional tunes that are difficult to name than for traditional tunes that are easy to name, even when these tunes are "new" (i.e., when they are recognition-test lures).

To determine if a labeling strategy were used by either the young adults or the elderly subjects, we analyzed the naming data from Experiments 2 and 3 to assess nameability of each of the 16 traditional tunes. Nameability scores for each of these tunes were derived as in Experiment 1 for both young-adult and elderly subjects. This allowed us to identify 6 traditional tunes with greater-than-average nameability for both young-adult and elderly subjects and 6 traditional tunes with lower-than-average nameability for both young-adult and elderly subjects (the remaining 4 tunes had high or low scores in one age group only). We computed hit and false-alarm rates in old/new recognition for high-nameable items and low-nameable items separately, collapsing across the first and second tests.⁵

Table 4 shows hit and false-alarm rates for traditional tunes of high and low nameability. Note that the high-nameable items evoked lower hit rates and lower false-alarm rates than did the low-nameable items, but only in the elderly group. The age \times nameability interaction was supported in an ANOVA of hit rates for high- and low-nameable items [$F(1,40) = 5.96, MS_e = 0.036$], as well as in a second ANOVA of false-alarm rates for these items [$F(1,40) =$

Table 4
Hit and False-Alarm Rates in Old/New Recognition
of High- and Low-Nameable Traditional Melodies
by Young-Adult and Elderly Subjects in Experiment 3

Melody Type	Young Adults (<i>n</i> =21)		Elderly Adults (<i>n</i> =21)		Difference
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Hit Rates					
High nameable	.98	.08	.75	.31	.23†
Low nameable	.93	.14	.91	.19	.02
False-Alarm Rates					
High nameable	.12	.25	.19	.27	-.07
Low nameable	.14	.27	.37	.32	-.23*

* $p < .05$, by *t* test. † $p < .01$, by *t* test.

4.22, $MS_e = 0.030$]. The former ANOVA also showed a main effect for age [$F(1,40) = 8.09$, $MS_e = 0.041$], and the latter ANOVA showed a main effect for nameability [$F(1,40) = 7.22$, $MS_e = 0.030$]. This pattern is evidence that elderly subjects, but not young adults, used a labeling strategy to set their criteria for old/new judgments in Experiment 3.⁶

What do these findings imply about the process of source monitoring in relationship to age? Although some studies have suggested that older individuals are impaired at source monitoring (Dywan & Jacoby, 1989; Bartlett et al., 1990), Experiment 3 supported no age difference of this type. At the same time, this experiment suggested that young and older subjects use different strategies for making old/new judgments to more and less familiar tunes. Elderly subjects appear to employ a labeling strategy, tightening criteria for tunes they can identify and/or loosening criteria for tunes they cannot (Figure 3, panel A). It is not yet clear what younger subjects do, but an attractive idea is that they employ a recollection strategy: decrementing familiarity for tunes that initially feel familiar but do not evoke retrieval of episodic information concerning the study episode (Figure 3, panel B).

GENERAL DISCUSSION

Most of the present findings are consistent with those of Bäckman (1991), and Bäckman and Herlitz (1990), despite the change in stimulus materials (tunes vs. faces and proper names). In agreement with the study by Bäckman and Herlitz (1990), Experiment 1 showed that AD patients are impaired relative to controls in old/new recognition of highly familiar items (traditional tunes or well-known faces) than in old/new recognition of less familiar items (novel tunes or lesser known faces). Moreover, the pattern of the deficit in both studies was that the AD patients were impaired at reducing false-alarm rates for the more familiar items. Whereas the healthy control subjects showed the mirror effect of lower false-alarm rates along with higher hit rates for more familiar items than for less familiar items, the AD patients did not.

Experiment 2 replicated Bäckman's (1991) findings that though old/new recognition is impaired in healthy elderly

subjects, relative to young adults, this age-related impairment was about the same size for more and less familiar items. Moreover, both the young-adult and the elderly subjects showed the mirror effect of higher hit rates and lower false-alarm rates for more familiar items than for less familiar items, just as in the Bäckman study. Hence, the qualitative pattern of the AD-related deficits was not reflected in the age-related deficits either here or in the study by Bäckman.

A discrepancy from the research by Bäckman was found in Experiment 3. Whereas Experiment 2 used a blocked-list design in which more and less familiar items were consigned to separate lists, Experiment 3 used a mixed-list design in which more and less familiar items were mixed in the study lists and tests. Bäckman (1991) used mixed-list procedures as well; however, he observed the mirror effect among both young and healthy elderly subjects, whereas Experiment 3 failed to produce the mirror effect in either age group. Although hit rates were higher for traditional tunes than for novel tunes, false-alarm rates were not generally lower for traditional tunes than for novel tunes. Indeed, the opposite effect was statistically significant in the second memory test.

Just how to consider this apparent discrepancy is difficult to judge for two reasons. First, the mirror effect was not statistically evaluated in the research by Bäckman and colleagues. Second, that research differed from the present investigation with respect to the nature of the stimuli (faces vs. tunes) and strength of the familiarity manipulation: whereas Bäckman and colleagues used stimuli that were more versus less familiar (i.e., dated vs. contemporary faces), we used stimuli that were well known versus entirely unknown (i.e., traditional vs. novel tunes). Prior studies using such manipulations (e.g., studies comparing common words with unknown words or "nonsense" letter strings) have not always supported the mirror effect. Of the eight such studies reviewed by Glanzer and Adams (1985), only four showed a mirror effect (see their Table 4). Hence, the mirror effect in recognition memory for well-known versus unknown stimuli appears to be not an invariant phenomenon but rather a pattern that sometimes is found, and sometimes is not, depending on conditions that are not well understood.

With due consideration to some discrepancies in method as well as in results, the present experiments converge with Bäckman's in supporting two conclusions. First, whereas AD-related deficits in recognition memory sometimes are more marked for well-known stimuli than for less known or unknown stimuli, age-related deficits in recognition memory appear similar in size for more and less known stimuli, whether these stimuli are faces or tunes. Second, whereas the mirror effect is less strongly supported among AD patients than among age-matched controls, healthy young adults and elderly persons do not differ systematically with respect to the mirror effect. Indeed, the present study shows that the mirror effect with well-known tunes versus unknown tunes is less related to age than to the design variable of mixed versus blocked presentation.

Turning to theoretical issues, Experiment 1 was consistent with evidence (Bäckman & Herlitz, 1990) that AD-related deficits in old/new recognition reflect underlying deficits in utilizing prior knowledge. In addition, the pattern of performance in the AD groups both here and in Bäckman and Herlitz suggests that AD-related deficits are linked to the use of global familiarity in making old/new judgments. The key finding here is that though AD patients showed a greater impairment in d' with well-known stimuli than with poorly known stimuli, this group \times tune type interaction was due entirely to false-alarm rates: The largest group difference was that AD subjects exceeded controls in false alarms to well-known stimuli. An interesting consequence of this pattern was that AD patients were more liberal in recognizing traditional tunes than in recognizing novel tunes, relative to control subjects. These findings from Experiment 1 converge with our initial pilot study in which several AD patients attempted old/new recognition of traditional and novel tunes using a mixed-list design. The patients simply called traditional tunes "old" and novel items "new" regardless of true "oldness." Like the findings of Experiment 1, this outcome makes sense if AD patients retrieve sufficient prior knowledge to perceive well-known items as more familiar than unknown items and make their recognition decisions in accordance with a unidimensional familiarity model (i.e., a model that assumes the use of unadjusted levels of global familiarity and a single, fixed, criterion). What control subjects do was clarified considerably in Experiments 2 and 3.

Experiment 2 was conducted to test the hypothesis that age-related deficits in recognition memory differ from AD-related deficits in such memory in involving simply a quantitative reduction in performance, with no departure from the mirror effect. Indeed, no hints of a departure from the mirror effect were found. In both age groups, we observed the mirror pattern of higher hit rates and lower false-alarm rates for traditional tunes than for novel tunes. We also observed that though d' scores were higher for traditional tunes than for novel tunes, c scores were approximately zero in both age groups and with both types of tunes, in conformity with the mirror effect as defined by Hintzman et al. (1994).

Taken by themselves, the findings of Experiment 2 might suggest that healthy young and older subjects use some variable other than global familiarity for making old/new judgments (see Figure 2). However, Experiment 3 addressed the possibility that (1) healthy subjects base old/new judgments on global familiarity, but that (2) they compensate for differences in preexperimental familiarity by making adjustments in recognition criteria or familiarity itself for different types of item. If an adjustment strategy is to be effective, it must entail a source-monitoring process by which subjects distinguish between preexperimental sources of global familiarity and intraexperimental episodic sources. On the basis of the outcome of this source-monitoring process, subjects compensate for item differences in preexperimental familiarity through adjustments to criterion or familiarity itself (see Figure 3).

To test the adjustment hypothesis, in Experiment 3, we moved to a mixed-list design in which the source-monitoring process would be substantially more difficult than in the blocked-list design of Experiment 2. The mirror effect was not found in this study, suggesting that both young and older subjects (1) relied at least partially on global familiarity in making old/new judgments and (2) showed the mirror effect in Experiment 2 through use of some type of adjustment strategy.

Taken together, the results of Experiments 2 and 3 suggest that the use of global familiarity is an age-invariant aspect of old/new recognition memory, at least with tune stimuli. They also suggest that the mirror effect that sometimes occurs with such stimuli depends on a source-monitoring process—that is, a process of attributing global familiarity to some combination of preexperimental versus intraexperimental (episodic) sources. Finally, the findings of these two experiments reveal that there are age-related deficits in old/new recognition of tunes and that young and elderly subjects make their "old" and "new" judgments in different ways. Specifically, elderly subjects in Experiment 3 apparently used their knowledge about tune labels to set their criteria (or adjust familiarity) for traditional tunes and novel tunes. Young-adult subjects did not show this result, which suggests that they made old/new judgments not through using a labeling strategy but in some other way.

The present experiments provide no direct evidence on what this "other way" was. However, a good deal of prior research on memory and aging suggests that young adults exceed older adults in episodic memory tasks requiring recollection of events in their temporal/spatial context (see, e.g., Light, 1991). On the basis of this research, an interesting idea to explore in the future is that young adults made downward adjustments to the perceived familiarity of test items if these items (1) were high in global familiarity, but (2) failed to evoke recollection of information concerning the prior study episode. Such a recollection strategy might have allowed young-adult subjects to hold down false alarms to traditional tunes, at least in the first test where hints of the mirror effect were observed.

In general, our findings are opposed to a continuum view of age-related and AD-related deficits in recognition memory for tunes. Early stage AD patients perform in accordance with a unidimensional familiarity model, as if they were basing old/new judgments on global familiarity and a fixed criterion for "old" judgments. Because of this simple recognition strategy, they are particularly at risk for false-alarm errors with "new" items that are high in preexperimental familiarity. Healthy young and elderly subjects do not use this simple strategy. Although they apparently base "old" judgments on global familiarity, they have some ability to adjust their criteria (or familiarity itself) to maximize hits and minimize false alarms with stimuli that are high in preexperimental familiarity. Elderly adults differ from young adults in showing somewhat lower accuracy in old/new recognition and in making more use of a labeling strategy to adjust criterion (or

familiarity) for well-known versus unknown tunes. In any event, a principal weakness of AD patients—that of failing to make the best use of prior knowledge—appears not to be a weakness of healthy older persons. Our data indicate that elderly persons capitalize upon their prior knowledge as much as young adults do.

REFERENCES

- BÄCKMAN, L. (1991). Recognition memory across the adult life span: The role of prior knowledge. *Memory & Cognition*, **19**, 63-71.
- BÄCKMAN, L., & HERLITZ, A. (1990). The relationship between prior knowledge and face recognition memory in normal aging and Alzheimer's disease. *Journal of Gerontology*, **45**, 94-100.
- BARTLETT, J. C., & LESLIE, J. E. (1986). Aging and memory for faces versus pictures of faces. *Memory & Cognition*, **14**, 371-381.
- 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.
- BEATTY, W. W., ZAVADIL, K. D., & BAILLY, R. C. (1988). Preserved musical skill in a severely demented patient. *International Journal of Clinical Neuropsychology*, **10**, 158-164.
- BECKER, J. T., & LOPEZ, O. L. (1992). Episodic memory in Alzheimer's disease: Breakdown of multiple memory processes. In L. Bäckman (Ed.), *Memory functioning in dementia* (pp. 27-43). Amsterdam: Elsevier.
- BROWN, J., LEWIS, V. J., & MONK, A. F. (1977). Memorability, word frequency, and negative recognition. *Quarterly Journal of Experimental Psychology*, **29**, 461-473.
- CLAIR, A. A., & BERNSTEIN, B. (1990). A preliminary study of music therapy programming for severely regressed persons with Alzheimer's-type dementia. *Journal of Applied Gerontology*, **9**, 299-311.
- CRONIN-GOLOMB, A., KEANE, M. M., KOKODIS, A., CORKIN, S., & GROWDON, J. H. (1992). Category knowledge in Alzheimer's disease: Normal organization and a general retrieval deficit. *Psychology & Aging*, **7**, 359-366.
- CRYSTAL, H. A., GOBER, E., & MASUR, D. (1989). Preservation of musical memory in Alzheimer's disease. *Journal of Neurology, Neurosurgery, & Psychiatry*, **52**, 1415-1416.
- DAMASIO, A., VAN HOESEN, G. W., & HYMAN, B. T. (1990). Selectivity of neuropathological change. In M. Schwartz (Ed.), *Modular deficits in Alzheimer-type dementia* (pp. 83-100). Cambridge, MA: MIT Press, Bradford Books.
- DUARA, R., LONDON, E. D., & RAPAPORT, S. I. (1985). Changes in structure and energy metabolism of the aging brain. In C. E. Finch & E. L. Schneider (Eds.), *Handbook of the biology of aging* (pp. 595-616). New York: Van Nostrand Reinhold.
- DYWAN, J., & JACOBY, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology & Aging*, **5**, 379-387.
- 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.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.
- GLANZER, M., & ADAMS, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, **13**, 8-20.
- GLANZER, M., ADAMS, J. K., & IVERSON, G. (1991). Forgetting and the mirror effect in recognition memory: Concentrating of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 81-93.
- GLANZER, M., ADAMS, J. K., IVERSON, G., & KIM, K. (1993). The regularities of recognition memory. *Psychological Review*, **100**, 546-567.
- HALPERN, A. R. (1984). The organization of memory for familiar songs. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 496-512.
- HALPERN, A. R., BARTLETT, J. C., & DOWLING, W. J. (1995). Aging and expertise in the perception of musical transpositions. *Psychology & Aging*, **10**, 325-342.
- HERLITZ, A., LIPINSKA, B., & BÄCKMAN, L. (1992). Utilization of cognitive support for episodic remembering in Alzheimer's disease. In L. Bäckman (Ed.), *Memory functioning in dementia* (pp. 73-96). Amsterdam: Elsevier.
- HINTZMAN, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 201-205.
- HINTZMAN, D. L., CAULTON, D. A., & CURRAN, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 275-289.
- JACOBY, L. L., KELLEY, C. M., & DYWAN, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.
- LIGHT, L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, **42**, 333-376.
- LIPINSKA, B., BÄCKMAN, L., & HERLITZ, A. (1992). When Greta Garbo is easier to remember than Stefan Edberg: Influences of prior knowledge on recognition memory in Alzheimer's disease. *Psychology & Aging*, **7**, 214-220.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MCKHANN, G., DRACHMAN, D., FOLSTEIN, M., KATZMAN, R., PRICE, D., & STADLAN, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services task force on Alzheimer's disease. *Neurology*, **34**, 939-944.
- MORROW, D. G., LEIRER, V. O., & ALTIERI, P. A. (1992). Aging, expertise and narrative processing. *Psychology & Aging*, **7**, 376-388.
- NORBERG, A., MELIN, E., & ASPLUND, K. (1986). Reactions to music, touch and object presentation in the final stage of dementia: An exploratory study. *International Journal of Nursing Studies*, **23**, 315-323.
- OLDEROG-MILLER, K. A., & SMITH, J. M. (1989). The influence of group singing therapy on the behavior of Alzheimer's disease patients. *Journal of Music Therapy*, **26**, 58-70.
- SALMON, D. P., SHIMAMURA, A. P., BUTTERS, N., & SMITH, S. (1988). Lexical and semantic priming deficits in patients with Alzheimer's disease. *Journal of Clinical & Experimental Neuropsychology*, **10**, 477-494.
- SMITH, G. H. (1986). A comparison of the effects of three treatment interventions on cognitive functioning of Alzheimer patients. *Music Therapy*, **6**, 41-56.

NOTES

1. Since the traditional-tunes test always preceded the novel-tunes test, the effects of tune type within either subject group are difficult to interpret. Hence, our discussion is focused on differences in the effects of tune type between the subject groups. Experiment 2 incorporated more thorough counterbalancing to assess tune-type effects within groups of healthy subjects.
2. All of the ANOVAs in Experiment 2 included the between-subject variables of test order (traditional-novel vs. novel-traditional) and musical experience (0-2 years vs. more than 2 years of training). The only reliable effect involving musical experience was an anomalous age \times musical experience \times test order interaction with naming scores [$F(1,48) = 4.50$, $MS_e = 16.3$]: Naming scores were slightly higher in the traditional-novel condition than in the novel-traditional condition, except that elderly low-experience subjects showed the opposite effect.
3. The ANOVAs also showed reliable interactions between age and test order (traditional-novel vs. novel-traditional) for hits [$F(1,48) = 4.37$, $MS_e = 439.5$], false alarms [$F(1,48) = 6.02$, $MS_e = 372.6$], and d' s [$F(1,48) = 11.7$, $MS_e = 1.51$], though not for c scores ($p > .10$). The pattern in each of the first three cases was that of an age-related deficit with the traditional-novel order, but not with the novel-traditional order (e.g., the d' s for young-adult and elderly subjects averaged 2.86 and 1.54, respectively, with the traditional-novel order, and 1.74 and 2.03, respectively, with the novel-traditional order). We do not attempt to explain this pattern, but note that the traditional-novel order showing an age-related deficit produced an AD-related deficit in Experiment 2.
4. All ANOVAs included the factor of musical experience (0-2 years of training vs. more than 2 years). The familiarity ratings made to novel tunes showed a reliable main effect of this variable [$F(1,38) = 8.16$,

$MS_e = .30$], as proportions of false "familiar" judgments averaged .02 among the more experienced subjects and .07 among the low-experience subjects. Turning to old/new recognition, the ANOVAs of hit rates, false-alarm rates, and d' 's showed reliable age \times experience interactions [$F(1,38) = 6.67, MS_e = 6.18; F(1,38) = 4.66, MS_e = 11.33$; and $F(1,38) = 7.63, MS_e = 3.70$, respectively]. There was an age-related deficit that was larger in the more musically experienced groups than in the less experienced groups. This is the reverse of a "compensation" pattern (see Morrow, Leirer, & Altieri, 1992).

5. Since all traditional tunes evoked near-100% rates of "familiar" judgments in the last phase of the study, even the low-nameable items can be assumed to be familiar. The 6 high-nameable tunes were "America the Beautiful," "Old MacDonald Had a Farm," "Rudolf the Red-Nosed Reindeer," "When the Saints Go Marching In," "The Star-Spangled Banner," and "Yankee Doodle" (scores for young-adult and elderly subjects averaged .96 and .94, respectively, with SD s of .02 and .02). The 6 low-nameable tunes were "For He's a Jolly Good Fellow," "Mary Had a Little Lamb," "Here Comes the Bride," "Deck the Halls with Boughs of Holly," "This Land is Your Land," and "Oh Susanna" (scores for young-adult and elderly subjects averaged .85 and .83, respectively, with SD s of .11 and .08).

6. It is pertinent to ask whether a nameability analysis of Experiment 2 would produce the same outcome as that of Experiment 3. The adjustment hypothesis developed in this article implies that the answer should be "no." Experiment 2 used a blocked-test procedure that would make it quite easy to reduce false-alarm rates for traditional tunes by setting relatively conservative criteria for recognizing such tunes; no attempt to name tunes would be needed. In fact, the false-alarm rates in Experiment 2 supported no effects involving nameability. The hit rates showed an age \times nameability interaction [$F(1,54) = 4.44, MS_e = .07$], but it was opposite in form to that found in Experiment 3 (Table 4). The elderly subjects in Experiment 2 showed lower (instead of higher) hit rates for the low-nameable items (.58) than for the high-nameable items (.70), whereas the young-adult subjects showed somewhat higher hit rates for low-nameable items (.78) than for high-nameable items (.69). This pattern supported only with hits does not appear to be due to an adjustment process.

(Manuscript received November 29, 1993;
revision accepted for publication December 5, 1994.)