*Psychology and Aging* (in press)

# Effects of Time of Day on Age-related Associative Deficits

Elizabeth A. Maylor

University of Warwick, UK

Stephen P. Badham

Nottingham Trent University, UK

Author Note

Elizabeth A. Maylor, Department of Psychology, University of Warwick, United Kingdom, and Stephen P. Badham, Department of Psychology, Nottingham Trent University, United Kingdom.

We are grateful to Moshe Naveh-Benjamin for providing the stimuli, and to Roslyn Keal, Laura Steel, Sophie Stuttard, and Katherine Tyler for help with data collection.

Correspondence concerning this article should be addressed to Elizabeth A. Maylor, Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom. Email: e.a.maylor@warwick.ac.uk

**Abstract**

Time of day is known to influence cognition differently across age groups, with young adults performing better later than earlier in the day and older adults showing the opposite pattern. Thus age-related deficits can be smaller when testing occurs in the morning compared with the afternoon/evening, particularly for tasks requiring executive/controlled/inhibitory processes. Stronger influences of time of day were therefore predicted on associative than on item recognition memory based on their differential requirements for demanding recollective (rather than familiarity) processes. In two experiments, participants were presented with unrelated word pairs and then tested on both item recognition (old/new item?) and associative recognition (intact/recombined pair?). In Experiment 1, young adults were tested either in the morning or in the evening; recognition memory was better when time of testing matched participants' morningness-eveningness preferences, and more so for associative than for item memory. In Experiment 2, young and older adults (evening and morning types, respectively) were tested both in the morning and in the evening; again, recognition memory was better at participants' preferred times of day, especially for associative memory. Consequently, age-related associative deficits varied considerably – indeed more than four-fold – from a nonsignificant 8% for testing in the morning to a substantial 35% for testing in the evening, suggesting that it is important to consider time of day effects in future studies of the associative deficit hypothesis.

*Keywords:* time of day, aging, recognition, item memory, associative memory

Effects of Time of Day on Age-related Associative Deficits

It is well established that individuals differ in their chronotype – their preferred or peak arousal time of day – and that cognitive performance is optimal when testing times are in synchrony with these preferences (see review by Schmidt, Collette, Cajochen, & Peigneux, 2007). Importantly, young and older adults have quite different chronotypes, with older adults tending to prefer the morning and young adults tending to prefer later times of the day (see Hasher, Goldstein, & May, 2005; Yoon, May, & Hasher, 2000, for reviews). Researchers have consequently observed variation in the magnitude of age-related cognitive deficits depending on whether time of testing was earlier (smaller deficits) rather than later (larger deficits) in the day (e.g., May, Hasher, & Stoltzfus, 1993; Rowe, Hasher, & Turcotte, 2009; West, Murphy, Armilio, Craik, & Stuss, 2002). The aim of the present study was to explore time of day effects on recognition memory in the context of Naveh-Benjamin's (2000) associative deficit hypothesis of aging.

Chronotype is commonly assessed by Horne and Östberg's (1976) Morningness-Eveningness Questionnaire (MEQ), composed of questions regarding sleep-wake habits, alertness, appetite, and self-ratings of intellectual and physical functioning across the day. The MEQ has been successfully validated against a range of physiological measures, as summarized by May, Hasher, and Foong (2005). High scores indicate "morning" types, low scores indicate "evening" types, and mid-range scores are classified as "neither". MEQ data reported by Yoon et al. (2000) from 1,538 adults aged 18-23 and 608 adults aged 60-75 revealed approximate percentages of morning, neither and evening types of 5%, 57% and 38%, respectively, for young adults; the corresponding percentages for older adults were 73%, 25% and 2% (see also Intons-Peterson, Rocchi, West, McLellan, & Hackney, 1998; May et al., 1993).

With relatively few young morning types or older evening types, we would expect lower performance for young participants tested in the morning compared to later in the day, but higher performance for older participants tested in the morning compared to later in the day. Indeed, this pattern has been found across a range of cognitive tasks, including visuospatial working memory (Rowe et al., 2009), explicit memory (May et al., 2005), list learning (Hasher, Chung, May, & Foong, 2002), sentence recognition (May et al., 1993), the Rey Auditory Verbal Learning Test (Lehmann, Marks, & Hanstock, 2013), the Attention Network Test (Knight & Mather, 2013), negative priming (Intons-Peterson et al., 1998), inhibitory control tasks (May, 1999; May & Hasher, 1998; Rowe, Valderrama, Hasher, & Lenartowicz, 2006; see also Anderson, Campbell, Amer, Grady, & Hasher, 2014), the Stroop task (Borella, Ludwig, Dirk, & de Ribaupierre, 2011), verbal fluency (Iskandar et al., 2016; see also Allen, Grabbe, McCarthy, Bush, & Wallace, 2008), and several neuropsychological tests administered to older adults (Walters & Lesk, 2015). The same pattern has also been observed in rats in a non-matching-to-sample test (Winocur & Hasher, 2004).

In contrast to the above, a small minority of studies have observed synchrony effects in older but not in young adults (e.g., Hogan et al., 2009). An example relevant to the present study because of its focus on recognition memory is that of Intons-Peterson, Rocchi, West, McLellan, and Hackney (1999). These authors used the DRM (Deese, 1959; Roediger & McDermott, 1995) paradigm to examine the effects of age and testing at optimal versus nonoptimal times of day on false memories. Young and older participants studied lists of words, each thematically related to a nonstudied (critical lure) word. At test, both age groups falsely recognized the critical lures at rates similar to studied words and much higher than nonstudied unrelated words. Correct recognition of studied words was unaffected by whether or not participants were tested at their preferred (optimal) time of day. This was also the case for false recognition of critical lures for young participants. However, for older participants, false recognition was greater when tested nonoptimally than optimally, assumed to be indicative of an over-reliance on gist/familiarity at their nonpreferred time of day.

Contrary to Intons-Peterson et al. (1999), an unpublished DRM study in our own laboratory did find significant time of day effects on false recognition in undergraduate students who were tested either in the morning or in the evening. Participants also completed the MEQ and although their scores were generally skewed toward eveningness rather than morningness, there was still some variation within each group in terms of MEQ scores. We therefore predicted that relationships between MEQ scores and memory would differ in sign depending on time of testing. As morning types have high scores and evening types have low scores, we expected a negative correlation between false recognition and MEQ score for those tested in the morning, but a positive correlation for those tested in the evening. This was indeed the case, with correlations of -.33 (morning) and .37 (evening), which differed significantly, $z = 2.13$, $p = .033$. (For correct recognition, the MEQ correlations did not differ between the morning and evening conditions.)[1] Thus, it seems that time of day effects on false memories can be demonstrated in both young and older adults.

Reviewers of the synchrony literature (e.g., Hasher et al., 2005; Schmidt et al., 2007; Yoon et al., 2000) have argued that not all cognitive tasks/processes are equally susceptible to time of day effects. Instead, it seems that tasks involving executive/controlled/inhibitory processes associated with prefrontal cortex activity are especially vulnerable to the influence of circadian variation (see also Anderson et al., 2014; Bennett, Petros, Johnson, & Ferraro, 2008; Borella et al., 2011; Hahn et al., 2012; Manly, Lewis, Robertson, Watson, & Datta, 2002; May & Hasher, 1998; Pica, Pierro, & Kruglanski, 2014). Thus, in the case of memory, performance is better at optimal than at nonoptimal times of day for explicit but not for implicit memory, the former relying on “conscious, deliberate efforts” and the latter on “automatic, unconscious responses” (p. 99, May et al., 2005; see also Yang, Hasher, &

Wilson, 2007). Similarly, false recognition of plausible foils is a likely consequence at nonoptimal times if less effortful familiarity/gist-based processes are employed rather than more controlled recollective processes (Intons-Peterson et al., 1999; May et al., 1993; Yoon, 1997; cf. Bodenhausen, 1990, on the use of schema-driven heuristics in problem-solving at nonoptimal times).

How, then, might all this apply to the associative deficit hypothesis of aging (Naveh-Benjamin, 2000)? In the standard paradigm, participants are presented with pairs of unrelated items, with instructions to remember both the items themselves and their pairings. This is followed by two recognition memory tests: (1) an item recognition test in which participants have to respond old/new to previously-presented/new individual items, and (2) an associative recognition test in which participants have to respond old/new to intact/recombined pairings. Typically, results show an interaction between age (young vs. older) and test (item vs. associative) such that the age-related deficit in performance is greater for associative than for item memory, and this has been observed across a wide range of stimuli (see Old & Naveh-Benjamin, 2008). In other words, older adults seem to have particular difficulty in forming associations between items relative to remembering the items themselves. One suggestion, based on dual-process models of memory (Yonelinas, 2002), is that automatic familiarity-based processes are sufficient to complete an item test, whereas controlled recollection-based processes are required to complete an associative test (as the items within intact and recombined pairings are equally familiar). The latter processes are more impaired than the former in older adults (Light, Prull, La Voie, & Healy, 2000), hence the age-related associative deficit (see also Castel & Craik, 2003; Naveh-Benjamin et al., 2009; Shing, Werkle-Bergner, Li, & Lindenberger, 2008).

Combining these suggestions leads to the following predictions: First, there should be stronger effects of time of day on associative memory than on item memory because of the associative memory test's greater reliance on demanding recollective processes, as opposed to undemanding familiarity processes. This was investigated in Experiment 1 with young adults. Second, if associative memory is indeed more influenced by time of day than item memory, then age-related associative deficits should vary, depending on time of testing, from small (testing in the morning) to large (testing in the evening). This was investigated in Experiment 2 with both young and older adults.

## Experiment 1

Our task of recognition memory for unrelated word pairs was closely based on studies by Naveh-Benjamin and colleagues (e.g., Naveh-Benjamin, 2000; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003; Old & Naveh-Benjamin, 2008). The procedure followed that of our unpublished DRM study in that we randomly allocated undergraduate students to be tested either in the morning or in the evening. Note that one difference between our DRM study and that of Intons-Peterson et al. (1999) that may be crucial is the time of the later testing session ("3 p.m. or later" in Intons-Peterson et al., 1999; between 18:00 and 21:00 in our unpublished study). Thus, in both this experiment and in Experiment 2, we used between 18:00 and 21:00 as the evening period to compare with a morning period of between 07:30 and 10:30. Participants completed the MEQ, which allowed us again to examine (1) correlations between MEQ scores and memory, and (2) comparisons between participants tested at optimal versus nonoptimal times of day.

### Method

**Participants.** Fifty-nine adults (29 male; 30 female) aged 18-26 years (*M* = 20.3, *SD* = 1.2) took part in the experiment, all of whom were undergraduate students at the University of Warwick, UK, and had English as their first language. The experiment was approved by the Psychology Department's Research Ethics Committee at the University of Warwick and written informed consent was obtained. Thirty participants were randomly assigned to the morning condition and were tested between 07:30 and 10:30 hours, and 29 were assigned to the evening condition and tested between 18:00 and 21:00 hours; MEQ scores did not differ between the two groups (*M* = 45.7, *SD* = 16.2, and *M* = 43.1, *SD* = 12.1, respectively, $t < 1$). On Horne and Östberg's (1976) classifications of "definitely morning", "moderately morning", "neither", "moderately evening" and "definitely evening" types, there were 2, 5, 13, 3, and 7 participants, respectively, in the morning condition, and 1, 3, 13, 7 and 5, respectively, in the evening condition. Thus, consistent with normative data from young college-aged adults (e.g., Chelminski, Ferraro, Petros, & Plaud, 1997; Intons-Peterson et al., 1998; May & Hasher, 1998; May et al., 1993; Yoon et al., 2000), there were more evening types (37.3%) than morning types (18.6%), with the largest group (44.1%) being neither.

**Materials.** The experimental stimuli were 80 high-frequency bisyllabic nouns of 4-8 letters in length (*M* = 6.0; *SD* = 1.1). The study list comprised 30 word pairs in which the two words in each pair were unrelated semantically, acoustically and visually (e.g., *corset-dragon*, *parade-vessel*). The order of these 30 word pairs was randomized three times to produce three different versions of the study list, with a third of participants assigned to each order. In the item recognition memory test, there were 20 old words from the study list and 20 new words, presented in a random order. In the associative recognition memory test, there were 10 old intact word pairs from the study list and 10 "recombined" word pairs with the left-hand word from one pair presented together with the right-hand word from another pair, again presented in a random order. Each word from the study list appeared once only in the test phase, either in the item test or in the associative test. There were three different versions of the two tests, with a third of participants assigned to each version: Each word from the study phase appeared as an old word on the item test in one version, as one word of an intact pair on the associative test in another version, and as one word of a recombined pair on the associative test in the third version. Combined with the three versions of the study list, there were nine possible study-test combinations, with six or seven participants assigned to each combination. All stimuli were presented in lower case in a large font size on a computer screen. Single words appeared in the center of the screen whereas word pairs appeared with one word to the left and one word to the right of the center, separated by a hyphen between them.

**Procedure.** Participants were tested individually. They were told that they would see a list of 30 word pairs presented sequentially on a computer screen at a rate of 4 s per pair and were explicitly instructed to remember the words and the associations between them for later memory tests. Following the study phase, there was a distractor task of counting backwards in threes from 300 for 60 s before the item and associative recognition tests were administered, test order being counterbalanced across participants. In the item memory test, 40 words (20 old; 20 new) were presented sequentially on the screen and participants were required to respond verbally (yes/no) according to whether or not they had seen each word in the study phase. In the associative memory test, 20 word pairs (10 intact; 10 recombined) were presented sequentially and participants were required to respond yes/no according to whether or not they had seen the words paired together in the study phase. There were no time limits imposed on responding in the test phase – once a verbal response was made, the experimenter initiated the next trial. Before starting the experiment, participants received practice at the task with a study list of five word pairs, followed by item and associative tests, each with four trials. None of the practice words was included in the main experiment. At the end of the session, participants completed the MEQ.

## Results and Discussion

For each memory test, performance was scored in terms of hit rates (i.e., yes responses to old/intact stimuli as a percentage) and false alarm rates (i.e., yes responses to new/recombined stimuli as a percentage). Corrected recognition measures of performance were then calculated as percent hits minus percent false alarms (cf. May et al., 1993; Naveh-Benjamin, 2000). The signal detection measure of sensitivity, d′, was also calculated from participants' hit and false alarm rates (Stanislaw & Todorov, 1999) but the patterns of results were identical in almost every case in the present experiment, and were identical in every case in Experiment 2. We therefore focus on reporting hits minus false alarms, but mention d′ in parentheses where the results differed.

In view of the wide variation in MEQ scores, with morning and evening types in both conditions, we first examined time of day effects by comparing the morning and evening groups in terms of correlations between performance and MEQ score. For item memory, there was a positive correlation (i.e., better memory for morning types) of .48 ($p$ = .008) for those tested in the morning, and a negative correlation (i.e., better memory for evening types) of -.13 ($p$ = .499) for those tested in the evening.[2] These correlations differed significantly, $z$ = 2.37, $p$ < .018 (although for d′ the corresponding difference was only marginally significant, $z$ = 1.72, $p$ = .085). For associative memory, the corresponding correlations were .41 ($p$ = .025) and -.29 ($p$ = .130), which also differed significantly from each other, $z$ = 2.66, $p$ = .008. Thus, performance on both item and associative memory tests was systematically influenced by the match between time of testing and morningness-eveningness score.

To directly compare these effects on item and associative memory, a difference measure was calculated for each participant, namely, item memory minus associative memory. The correlation between this measure and MEQ score was negative (i.e., a smaller item-associative difference for morning types) for those tested in the morning (-.25) and positive (i.e., a smaller item-associative difference for evening types) for those tested in the evening (.25). Although neither correlation was significant ($p$ = .177 and .195, respectively), the difference between them fell just short of significance, $z = 1.86$, $p = .063$, indicating a tendency – in line with our predictions – for associative deficits to be smaller when tested at optimal rather than at nonoptimal times of day. (The corresponding correlations for d′ were -.24 and .31, which in this case differed significantly, $z = 2.04$, $p = .041$.)

As difference measures can be unreliable, our second approach was to divide participants into groups on the basis of their MEQ scores. To ensure adequate cell sizes, “morning” and “neither” types were pooled (cf. Matchock & Mordkoff, 2009) and compared with “evening” types. This resulted in 20 morning/neither and 10 evening types who were tested in the morning, and 17 morning/neither and 12 evening types who were tested in the evening. From the means in Figure 1, it can be seen that performance was higher when time of testing synchronized with MEQ type and lower when it did not, and this was especially the case for associative memory. These observations were confirmed by a repeated measures ANOVA with time of testing (morning vs. evening) and MEQ type (morning/neither vs. evening) as between-subjects factors, and memory type (item vs. associative) as the within-subjects factor. Item memory was significantly higher than associative memory, $F(1, 55) = 4.69$, $MSE = 310.48$, $p = .035$, $\eta_p^2 = .079$. There were no overall effects of time of testing, $F < 1$, or MEQ type, $F(1, 55) = 1.72$, $MSE = 1187.41$, $p = .195$, $\eta_p^2 = .030$, but there was a significant interaction between them, $F(1, 55) = 9.32$, $MSE = 1187.41$, $p = .003$, $\eta_p^2 = .145$. Importantly, this was further modified by a significant three-way interaction, $F(1, 55) = 5.42$, $MSE = 310.48$, $p = .024$, $\eta_p^2 = .090$. Follow-up two-way (time of testing by MEQ type) ANOVAs on item and associative memory separately revealed no effects of time of testing, $F < 1$, or of MEQ type, $p > .146$, but significant interactions between them for both item memory, $F(1, 55) = 5.44$, $MSE = 378.95$, $p = .023$, $\eta_p^2 = .090$, and associative memory, $F(1, 55) = 9.55$, $MSE = 1118.94$, $p = .003$, $\eta_p^2 = .148$. (For d′, the interaction for item memory only approached significance, $p = .063$.)

In summary, recognition memory was clearly affected by time of day such that performance was higher for participants tested at their optimal time (as indicated by their MEQ scores) than for participants tested at their nonoptimal time. Moreover, these synchrony effects were significantly larger for associative memory than for item memory.[3] To explore further the source of these effects, separate ANOVAs were conducted on hit rates and false alarm rates (see Table 1 for means). For hit rates, the only significant effect was the interaction between time of testing and MEQ type, $F(1, 55) = 4.85$, $MSE = 485.49$, $p = .032$, $\eta_p^2 = .081$, indicating that hit rate was higher for participants tested at their optimal time than for those tested at their nonoptimal time (76.8% vs. 70.2%, respectively, for item memory; 81.9% vs. 69.9% for associative memory). For false alarm rates, there was a significant effect of memory type, $F(1, 55) = 25.03$, $MSE = 102.99$, $p < .001$, $\eta_p^2 = .313$, a significant interaction between time of testing and MEQ type, $F(1, 55) = 8.11$, $MSE = 395.82$, $p = .006$, $\eta_p^2 = .129$, and a significant three-way interaction between time of testing, MEQ type and memory type, $F(1, 55) = 6.85$, $MSE = 102.99$, $p = .011$, $\eta_p^2 = .111$, indicating that false alarm rates were higher for participants tested at their nonoptimal time than for those tested at their optimal time, but reliably more so for associative memory (31.3% vs. 15.4%, respectively) than for item memory (16.5% vs. 10.8%). Thus, there were significant effects of synchrony in the present experiment for both hits and false alarms. However, the differential effect of synchrony on item versus associative memory was attributable more to false alarm rates than to hit rates.

## Experiment 2

Having established that recognition memory in young adults shows larger time of day effects for associative than for item memory, we turn now to exploring the implications for age-related associative deficits (Naveh-Benjamin, 2000). Young and older adults were compared in Experiment 2 using a design similar to several previous studies of memory (e.g., Hasher et al., 2002; May et al., 1993) in which participants were selected on the basis of their MEQ scores such that only young evening types and older morning types were tested. In this experiment, time of testing (morning/evening) was manipulated within- rather than between-subjects.

**Method**

**Participants.** Twenty-four young adults (16 male; 8 female) aged 19-23 years ($M$ = 21.0, $SD$ = 1.2) and 24 older adults (15 female; 9 male) aged 65-85 years ($M$ = 75.3, $SD$ = 6.3) took part in the experiment.[4] The young adults had more years of formal education ($M$ = 15.4, $SD$ = 1.5) than did the older adults ($M$ = 13.0, $SD$ = 4.0), $t(29.5) = 2.75$, $p = .010$, and rated their health better on a scale from 1 ("very poor") to 5 ("very good") (young $M$ = 4.6, $SD$ = 0.6; older $M$ = 3.7, $SD$ = 1.1), $t(34.6) = 3.39$, $p = .002$. Participants were all living independently in the local community and were free of any diagnosis of dementia or vision/speech/language difficulties. All were volunteers and had English as their first language. The experiment was approved by the Psychology Department's Research Ethics Committee at the University of Warwick and written informed consent was obtained.

Participants were recruited for the experiment according to their MEQ score to ensure that all young adults were evening types and all older adults were morning types on Horne and Östberg's (1976) classification. An additional 12 young adults were screened on the MEQ but not tested in the experiment because they were "neither" types; there were no additional older adults as the first 24 screened were all morning types. The 24 young participants comprised 1 "definitely evening" type and 23 "moderately evening" types; the 24 older participants comprised 4 "definitely morning" types and 20 "moderately morning" types. Mean MEQ scores were 38.0 ($SD$ = 3.0) and 64.8 ($SD$ = 5.6) for the young and older participants, respectively.

Each participant was tested twice, once in the morning between 07:30 and 10:30 hours and once in the evening between 18:00 and 21:00 hours. The start times of the morning tests did not differ significantly between age groups (*M* = 09:30 and 09:47 for young and older adults, respectively), $t(46) = 1.64$, $p = .107$. However, the evening tests started significantly later for young adults (*M* = 19:36) than for older adults (*M* = 19:03), $t(46) = 2.38$, $p = .021$. The order of the morning and evening tests was counterbalanced across participants in each age group. There was always at least one night's sleep separating the two tests, with the majority of participants having 2-4 nights of sleep between tests (*M* = 3.0 for both age groups).

**Materials.** The experimental stimuli were similar to those described in Experiment 1 (high-frequency bisyllabic nouns of 4-10 letters, with an average length of 6.2) except for the following changes. First, instead of just one study list (with accompanying item and associative tests), there were six different study lists. Participants saw one study list in the morning session and another study list in the evening session. Lists were counterbalanced such that each was used equally often across the two age groups and two times of testing. Second, study list length was increased from 30 word pairs to 34 word pairs to include buffer pairs, two at the beginning and two at the end of each list; memory for these buffer pairs was not tested.

**Procedure.** This was identical to Experiment 1 except for the following changes. First, the MEQ was administered prior to the first of the two sessions to allow the early termination of the experiment for volunteers who did not meet our criteria (see Participants section). Second, participants were tested both in the morning and in the evening (order counterbalanced as already noted). The order of the item and associative tests was counterbalanced across age groups and times of testing; however, for each individual participant, the item-associative order was kept the same across the two sessions.

**Results and Discussion**

Figure 2 displays the mean percentages for corrected recognition (hits minus false alarms) for item and associative memory as a function of time of testing (morning vs. evening) for each age group. First, as expected, older adults' memory was poorer than that of young adults, and this age-related deficit was greater for associative than for item memory, consistent with the associative deficit hypothesis (Naveh-Benjamin, 2000).[5] Second, the data pattern for the young adults, who were all evening types, successfully replicated that of the young evening types in Experiment 1 (compare the right-hand bars of Figures 1 and 2). Thus, whether time of testing was manipulated between subjects (Experiment 1) or within subjects (present experiment), performance was poorer in the morning than in the evening, the difference being more evident for associative than for item memory. Third, the older adults (who were all morning types) performed better in the morning than in the evening for associative memory, with little effect of time of testing on item memory. Note that the differential effects of synchrony on item versus associative memory were similar in magnitude for young and older adults, with item-associative differences larger at nonoptimal than at optimal times by 13.8% (young) and 12.9% (older).

A repeated measures ANOVA was conducted on these data with age group (young vs. older) as the between-subjects factor, and time of testing (morning vs. evening) and memory type (item vs. associative) as within-subjects factors. There were significant effects of age group, $F(1, 46) = 31.15$, $MSE = 1113.04$, $p < .001$, $\eta_p^2 = .404$, and of memory type, $F(1, 46) = 54.86$, $MSE = 342.75$, $p < .001$, $\eta_p^2 = .544$, with an interaction between them, $F(1, 46) = 15.81$, $MSE = 342.75$, $p < .001$, $\eta_p^2 = .256$. The effect of time of testing approached significance, $F(1, 46) = 3.78$, $MSE = 291.89$, $p = .058$, $\eta_p^2 = .076$, and time of testing interacted significantly with age group, $F(1, 46) = 10.31$, $MSE = 291.89$, $p = .002$, $\eta_p^2 = .183$. Finally, the three-way interaction was also significant, $F(1, 46) = 6.68$, $MSE = 319.61$,

$p$ = .013, $\eta_p^2$ = .127, confirming that synchrony effects (i.e., young adults better in the evening than in the morning; older adults better in the morning than in the evening) were more apparent for associative than for item memory. Indeed, separate two-way (age group by time of testing) ANOVAs on item and associative memory revealed significant effects of age group in both cases, $p < .001$, and no significant effects of time of testing in both cases, $p > .158$, but whereas there was no interaction for item memory, $F < 1$, the interaction for associative memory was significant, $F(1, 46) = 13.98$, $MSE$ = 365.13, $p$ = .001, $\eta_p^2$ = .233.

As before, separate three-way ANOVAs were conducted on the hit and false alarm rates (see Table 2 for means). First, hit rates were higher for young adults than for older adults, $F(1, 46) = 6.11$, $MSE$ = 802.33, $p$ = .017, $\eta_p^2$ = .117, with the reverse for false alarm rates, $F(1, 46) = 22.10$, $MSE$ = 611.03, $p < .001$, $\eta_p^2$ = .324. Second, again as expected, the age group by memory type interaction was not significant for hits, $F < 1$, but was significant for false alarms, $F(1, 46) = 17.35$, $MSE$ = 237.75, $p < .001$, $\eta_p^2$ = .274. Third, for hit rates, the interaction between age group and time of testing was not significant, $F(1, 46) = 2.13$, $MSE$ = 191.49, $p$ = .151, $\eta_p^2$ = .040, but the three-way interaction was significant, $F(1, 46) = 5.12$, $MSE$ = 215.13, $p$ = .028, $\eta_p^2$ = .100. Thus hit rates were higher when participants were tested at their optimal time than when tested at their nonoptimal time, but only for associative memory (69.2% vs. 61.5%) and not for item memory (64.7% vs. 66.6%). For false alarm rates, there was a significant interaction between age group and time of testing, $F(1, 46) = 7.29$, $MSE$ = 164.63, $p$ = .010, $\eta_p^2$ = .137, but no three-way interaction, $F(1, 46) = 1.32$, $MSE$ = 127.67, $p$ = .256, $\eta_p^2$ = .028. False alarm rates were higher at nonoptimal than optimal times for both item memory (15.1% vs. 12.0%) and associative memory (36.5% vs. 29.6%). However, note that false alarm rates for associative memory in older adults at their nonoptimal time were close to chance levels, limiting the possibility of a three-way interaction.

Similarly, a potential concern with the data in Figure 2 is the low performance of the older group for associative memory, which was quite close to floor especially in the evening. To examine this further, data were excluded from participants whose overall hit rates did not exceed their overall false alarm rates for associative memory. This applied to five out of the 24 older adults and so the five poorest performing young participants on the associative tests were also excluded to retain comparability between groups (cf. Naveh-Benjamin et al., 2009). The original three-way ANOVA was repeated on this reduced data set, with an identical pattern of results (including the crucial two- and three-way interactions) except that the previously marginal main effect of time of testing no longer approached significance, $F < 1$. This time, the differential effects of synchrony on item versus associative memory were identical for young and older adults, with item-associative differences larger at nonoptimal than at optimal times by 13.9% for both age groups.

Finally, in view of significant influences of synchrony on associative memory but not on item memory, it is conceivable that age-related associative deficits might be eliminated under certain combinations of testing young and older adults at optimal and nonoptimal times. To test this possibility, four ANOVAs were conducted on the present data with age group (between-subjects) and memory type (within-subjects) as factors, comparing the following: young (morning) versus older (morning); young (morning) versus older (evening); young (evening) versus older (morning); young (evening) versus older (evening). In all cases, there were significant effects of both age group and memory type. However, whereas the interaction was nonsignificant in the first case, it was significant in the remaining three cases, $p = .329$, .008, .005, and < .001, respectively ($\eta_p^2 = .021$, .142, .161, and .361, respectively). Figure 3 displays the amounts by which item-associative differences of older participants exceeded those of young participants (i.e., age-related associative deficits) for each of these four age group × time of testing combinations. Note that the age-related associative deficit ranged from small (and nonsignificant) for young adults at their nonoptimal time and older adults at their optimal time, to moderate (and significant) for young and older adults both at their nonoptimal times or both at their optimal times, and finally to large (and significant) for young adults at their optimal time and older adults at their nonoptimal time.

## General Discussion

To summarize, we first predicted that there would be stronger influences of time of day on associative memory than on item memory in a recognition task because of their differential demands on the type of frontal processes thought to be susceptible to synchrony effects. In Experiment 1, young adults showed better recognition memory when time of testing matched rather than mismatched their morningness-eveningness preferences, and this was indeed reliably more apparent for associative than for item memory. This novel finding was successfully replicated by the young adults in Experiment 2, employing a different design. The inclusion of older adults in Experiment 2 revealed that they similarly showed time of day effects that were greater for associative than for item memory. As a result, consistent with our second prediction, the age-related associative deficit (i.e., the extent to which the age deficit for item memory was exceeded by that of the associative deficit) varied hugely from small and nonsignificant (8% for testing in the morning)[6] to large and significant (35% for testing in the evening).

In the present experiments, time of day effects were evident in terms of both lower hit rates and higher false alarm rates at nonoptimal compared to optimal times, although the effects were numerically larger overall for false alarms than for hits. This aligns well with previous recognition memory literature in which both young and older adults have shown time of day effects either on both hits and false alarms (young: May et al., 1993; older: Ryan, Hatfield, & Hofstetter, 2002, and Yoon, 1997), or on false alarms only (young: our unpublished DRM study, and Yoon, 1997; older: Intons-Peterson et al., 1999, and May et al., 1993). Notably, the time of day effects in the present study were always larger for associative than for item memory for both hits and false alarms – significantly so for false alarms in Experiment 1 and for hits in Experiment 2.

In terms of cognitive processes thought to be particularly susceptible to time of day effects (see earlier), it could be argued that at nonoptimal times, participants might fail to carry out the strategic/elaborative encoding required to form strong links between items in a pair, hence reduced hit rates to intact pairs. Also, as previously discussed, increased false alarm rates to recombined pairs would be expected if participants at nonoptimal times responded more on the basis of familiarity rather than engaging in recollective processing or strategic retrieval (see Cohn, Emrich, & Moscovitch, 2008). Note that this should not be confused with a shift to a more liberal response criterion at nonoptimal times of day. Indeed, we calculated the response bias measure, c (Stanislaw & Todorov, 1999), for which positive values indicate a conservative bias (i.e., toward responding no/not seen before) and negative values indicate a liberal bias (i.e., toward responding yes/seen before), and found that the only significant effect in our experiments was that of memory type: Responding was more conservative for item memory than for associative memory (Experiment 1: $c_{item}$ = 0.25 [95% confidence interval = 0.15 – 0.35] and $c_{associative}$ = 0.01 [-0.08 – 0.10]; Experiment 2: $c_{item}$ = 0.39 [0.30 – 0.49] and $c_{associative}$ = 0.03 [-0.09 – 0.15]). This replicates a similar difference in response criterion between item and associative memory observed in a study by Bender, Naveh-Benjamin, and Raz (2010). We found no other main effects or interactions including, crucially, no interactions between MEQ type/age group and time of testing. Thus there was no evidence in the present study for any influence of time of day on response criterion.

We have so far discussed our findings in terms of particular cognitive processes thought to be especially vulnerable to time of day effects as indicated by the literature reviewed in the introduction. It should be acknowledged, however, that there are alternative frameworks to "process" dissociation explanations for recognition memory. Benjamin's (2010) DRYAD theory (see also Benjamin, Diaz, Matzen, & Johnson, 2012), for example, argues instead for a "representational" account that explains why any variable that compromises memory (in our case, nonoptimal time of day) should selectively impair context (here, associative) memory more than content (here, item) memory. The present data do not allow us to distinguish between these various accounts. But whatever the actual mechanism, the practical outcome at least is clear, namely, age-related associative deficits can vary widely depending on the particular combination of time of testing and age group as illustrated by Figure 3.

Of course, it remains an empirical question whether such a wide variation in effect sizes would occur with testing times more typically employed in laboratory studies of cognitive aging (recall that the morning and evening testing times in Experiment 2 were, on average, 09:39 and 19:20, respectively). However, an increasing number of experiments are being conducted online (e.g., Colloff, Wade, Wixted, & Maylor, 2017), possibly by volunteers taking part either before or after a day's work. The general importance of taking the time of day factor into account in the design of cognitive aging studies has already been discussed elsewhere (e.g., Hasher et al., 2005; May et al., 1993). Note that it may be more of an issue with respect to older participants because of their much stronger preference for morning than young participants' preference for afternoon/evening (e.g., Yoon et al., 2000). This may partly explain why there are some studies that find time of day effects in older but not in young adults (see earlier), particularly those that do not collect MEQ scores but simply allocate young and older participants at random to morning/afternoon testing sessions (e.g., Borella et al., 2011). Note that had we not taken account of MEQ scores in our analyses of Experiment 1, we would have found no significant differences in memory between the morning and evening conditions. (In contrast, of course, the solely morning-type older participants in Experiment 2 were actually the first 24 older adults recruited for that study.)

One possible limitation of the present study is that only the MEQ was employed to measure individual circadian rhythms; although MEQ scores do correlate well with circadian fluctuations in body temperature and arousal (Horne & Östberg, 1976; Schmidt et al., 2007), future work should include other measures such as sleep activity. This work could also be extended to cued recall for which age-related associative deficits are perhaps more evident than for recognition memory (see Old & Naveh-Benjamin, 2008).

**References**

Allen, P. A., Grabbe, J., McCarthy, A., Bush, A. H., & Wallace, B. (2008). The early bird does not get the worm: Time-of-day effects on college students' basic cognitive processing. *American Journal of Psychology, 121,* 551-564.

Anderson, J. A. E., Campbell, K. L., Amer, T., Grady, C. L., & Hasher, L. (2014). Timing is everything: Age differences in the cognitive control network are modulated by time of day. *Psychology and Aging, 29,* 648-657.

Badham, S. P., & Maylor, E. A. (2011). Age-related associative deficits are absent with nonwords. *Psychology and Aging, 26,* 689-694.

Bender, A. R., Naveh-Benjamin, M., & Raz, N. (2010). Associative deficit in recognition memory in a lifespan sample of healthy adults. *Psychology and Aging, 25,* 940-948.

Bennett, C. L., Petros, T. V., Johnson, M., & Ferraro, F. R. (2008). Individual differences in the influence of time of day on executive functions. *American Journal of Psychology, 121,* 349-361.

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review, 117,* 1055-1079.

Benjamin, A. S., Diaz, M., Matzen, L. E., & Johnson, B. (2012). Tests of the DRYAD theor of the age-related deficit in memory for context: Not about context, and not about aging. *Psychology and Aging, 27,* 418-428.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science, 1,* 319-322.

Borella, E., Ludwig, C., Dirk, J., & de Ribaupierre, A. (2011). The influence of time of testing on interference, working memory, processing speed, and vocabulary: Age differences in adulthood. *Experimental Aging Research, 37,* 76-107.

Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging, 18,* 873-885.

Chelminski, I., Ferraro, F. R., Petros, T., & Plaud, J. J. (1997). Horne and Östberg questionnaire: A score distribution in a large sample of young adults. *Personality and Individual Differences, 23,* 647-652.

Cohn, M., Emrich, S. M., & Moscovitch, M. (2008). Age-related deficits in associative memory: The influence of impaired strategic retrieval. *Psychology and Aging, 23,* 93-103.

Colloff, M. F., Wade, K. A., Wixted, J. T., & Maylor, E. A. (2017). A signal-detection analysis of eyewitness identification across the adult lifespan. *Psychology and Aging, 32,* 243-258.

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

Hahn, C., Cowell, J. M., Wiprzycka, U. J., Goldstein, D., Ralph, M., Hasher, L., & Zelazo, P. D. (2012). Circadian rhythms in executive function during the transition to adolescence: The effect of synchrony between chronotype and time of day. *Developmental Science, 15,* 408-416.

Hasher, L., Chung, C., May, C. P., & Foong, N. (2002). Age, time of testing, and proactive interference. *Canadian Journal of Experimental Psychology, 56,* 200-207.

Hasher, L., Goldstein, D., & May, C. P. (2005). It’s about time: Circadian rhythms, memory, and aging. In C. Izawa & N. Ohta (Eds.), *Human learning and memory: Advances in theory and application* (pp. 199-217). Mahwah, NJ: Lawrence Erlbaum Associates.

Hogan, M. J., Kelly, C. A. M., Verrier, D., Newell, J., Hasher, L., & Robertson, I. H. (2009). Optimal time-of-day and consolidation of learning in younger and older adults. *Experimental Aging Research, 35,* 107-128.

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4,* 97-110.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 362-376.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1999). Aging, testing at preferred or nonpreferred times (testing optimality), and false memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 23-40.

Iskandar, S., Murphy, K. J., Baird, A. D., West, R., Armilio, M., Craik, F. I. M., & Stuss, D. T. (2016). Interacting effects of age and time of day on verbal fluency performance and intraindividual variability. *Aging, Neuropsychology, and Cognition, 23,* 1-17.

Knight, K., & Mather, M. (2013). Look out – it's your off-peak time of day! Time of day matters more for alerting than for orienting or executive attention. *Experimental Aging Research, 39,* 305-332.

Lehmann, C. A., Marks, A. D. G., & Hanstock, T. L. (2013). Age and synchrony effects in performance on the Rey Auditory Verbal Learning Test. *International Psychogeriatrics, 25,* 657-665.

Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238-300). Oxford: Oxford University Press.

Manly, T., Lewis, G. H., Robertson, I. H., Watson, P. C., & Datta, A. K. (2002). Coffee in the cornflakes: Time-of-day as a modulator of executive response control. *Neuropsychologia, 40,* 1-6.

Matchock, R. L., & Mordkoff, J. T. (2009). Chronotype and time-of-day influences on the alerting, orienting, and executive components of attention. *Experimental Brain Research, 192,* 189-198.

May, C. P. (1999). Synchrony effects in cognition: The costs and a benefit. *Psychonomic Bulletin & Review, 6,* 142-147.

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 363-379.

May, C. P., Hasher, L., & Foong, N. (2005). Implicit memory, age, and time of day: Paradoxical priming effects. *Psychological Science, 16,* 96-100.

May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science, 4,* 326-330.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1170-1187.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 826-837.

Naveh-Benjamin, M., Shing, Y. L., Kilb, A., Werkle-Bergner, M., Lindenberger, U., & Li, S-C. (2009). Adult age differences in memory for name-face associations: The effects of intentional and incidental learning. *Memory, 17,* 220-232.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23,* 104-118.

Pica, G., Pierro, A., & Kruglanski, A. W. (2014). Effect of circadian rhythms on retrieval-induced forgetting. *Cognitive Processing, 15,* 29-38.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803-814.

Rowe, G., Hasher, L., & Turcotte, J. (2009). Age and synchrony effects in visuospatial working memory. *The Quarterly Journal of Experimental Psychology, 62,* 1873-1880.

Rowe, G., Valderrama, S., Hasher, L., & Lenartowicz, A. (2006). Attentional disregulation: A benefit for implicit memory. *Psychology and Aging, 21,* 826-830.

Ryan, L., Hatfield, C., & Hofstetter, M. (2002). Caffeine reduces time-of-day effects on memory performane in older adults. *Psychological Science, 13,* 68-71.

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology, 24,* 755-789.

Shing, Y. L., Werkle-Bergner, M., Li, S-C., & Lindenberger, U. (2008). Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General, 137,* 495-513.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31,* 137-149.

Walters, E. R., & Lesk, V. E. (2015). Time of day and caffeine influence some neuropsychological tests in the elderly. *Psychological Assessment, 27,* 161-168.

West, R., Murphy, K. J., Armilio, M. L., Craik, F. I. M., & Stuss, D. T. (2002). Effects of time of day on age differences in working memory. *Journal of Gerontology: Psychological Sciences, 57B,* P3-P10.

Winocur, G., & Hasher, L. (2004). Age and time-of-day effects on learning and memory in a non-matching-to-sample test. *Neurobiology of Aging, 25,* 1107-1115.

Yang, L., Hasher, L., & Wilson, D. E. (2007). Synchrony effects in automatic and controlled retrieval. *Psychonomic Bulletin & Review, 14,* 51-56.

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

Yoon, C. (1997). Age differences in consumers' processing strategies: An investigation of moderating influences. *Journal of Consumer Research, 24,* 329-342.

Yoon, C., May, C. P., & Hasher, L. (2000). Aging, circadian arousal patterns, and cognition. In D. C. Park & N. Schwarz (Eds.), Cognitive aging: A primer (pp. 151-171). Hove, East Sussex: Psychology Press.

**Footnotes**

[1]Adopting Intons-Peterson et al.'s (1999) approach, we additionally analyzed a subset of our data, comparing those who were tested at their optimal and nonoptimal times (i.e., excluding those who were neither morning nor evening MEQ types). Again, false (but not correct) recognition rates differed significantly between these groups, being higher at nonoptimal times.

[2]The asymmetry in these correlations can be attributed to participants with higher MEQ scores (i.e., morning types) performing slightly better overall than participants with lower MEQ scores (i.e., evening types).

[3]The same conclusions were reached from analyses restricted to morning and evening types only, tested at optimal ($n$ = 19) and nonoptimal times ($n$ = 14). In a two-way mixed ANOVA, there was a marginal effect of memory type, $F(1, 31) = 3.25$, $MSE = 226.35$, $p = .081$, $\eta_p^2 = .095$, a significant effect of optimality, $F(1, 31) = 4.83$, $MSE = 1222.59$, $p = .036$, $\eta_p^2 = .135$, and a significant interaction between them, $F(1, 31) = 5.27$, $MSE = 226.35$, $p = .029$, $\eta_p^2 = .145$. Mean recognition performance for those tested at optimal times was 64.5% for item memory and 66.3% for associative memory; for those tested at nonoptimal times, the means were 53.9% (item) and 38.6% (associative).

[4]Data from this experiment were included – but without the time of testing factor – in a study by Badham and Maylor (2011; words condition).

[5]In view of differences between the two age groups on background variables (gender, education and self-rated health), we conducted a number of analyses to explore their possible influences on this age-related associative deficit and found no significant effects of any of these factors.

[6]It should be emphasized that although the age-related associative deficit was reduced in magnitude to nonsignificance when both young and older adults were tested in

the morning, the overall age-related memory deficit nevertheless remained significant and substantial.

Table 1

*Means (and Standard Deviations) for Hit and False Alarm Rates (%) in Experiment 1 as a Function of Type of Memory (Item Vs. Associative), Time of Test (Morning Vs. Evening), and Morningness-Eveningness Questionnaire (MEQ) Type (Morning/Neither Vs. Evening)*

| | | MEQ Type | | | |
|---|---|---|---|---|---|
| | | Morning/Neither | | Evening | |
| Memory | Time of test | Hits | False Alarms | Hits | False Alarms |
| Item | | | | | |
| | Morning | 79.8 (12.4) | 8.3 (8.8) | 69.0 (14.7) | 17.5 (11.6) |
| | Evening | 71.5 (19.3) | 15.6 (15.8) | 73.8 (13.2) | 13.3 (7.8) |
| Associative | | | | | |
| | Morning | 83.0 (24.1) | 12.5 (15.5) | 68.0 (18.1) | 35.0 (22.2) |
| | Evening | 71.8 (18.1) | 27.6 (23.3) | 80.8 (16.8) | 18.3 (14.7) |

Table 2

*Means (and Standard Deviations) for Hit and False Alarm Rates (%) in Experiment 2 as a Function of Type of Memory (Item Vs. Associative), Time of Test (Morning Vs. Evening), and Age Group (Older Vs. Young; Morning and Evening Types, Respectively)*

| | | Age Group | | | |
|---|---|---|---|---|---|
| | | Older | | Young | |
| Memory | Time of test | Hits | False Alarms | Hits | False Alarms |
| Item | | | | | |
| | Morning | 57.9 (17.9) | 15.6 (12.3) | 68.5 (17.5) | 11.3 (10.1) |
| | Evening | 64.6 (14.9) | 19.0 (14.1) | 71.5 (17.8) | 8.3 (11.2) |
| Associative | | | | | |
| | Morning | 62.5 (21.5) | 44.2 (24.5) | 66.3 (21.8) | 25.0 (19.1) |
| | Evening | 56.7 (18.8) | 47.9 (23.4) | 75.8 (19.5) | 15.0 (13.8) |

*Figure 1.* Recognition performance in Experiment 1 as percentages of hits minus false alarms for item memory (top panel) and associative memory (bottom panel) for participants classified as morning/neither and evening types on the Morningness-Eveningness Questionnaire (MEQ), tested in the morning (light bars) or in the evening (dark bars). Error bars represent ±1 *SEM*.

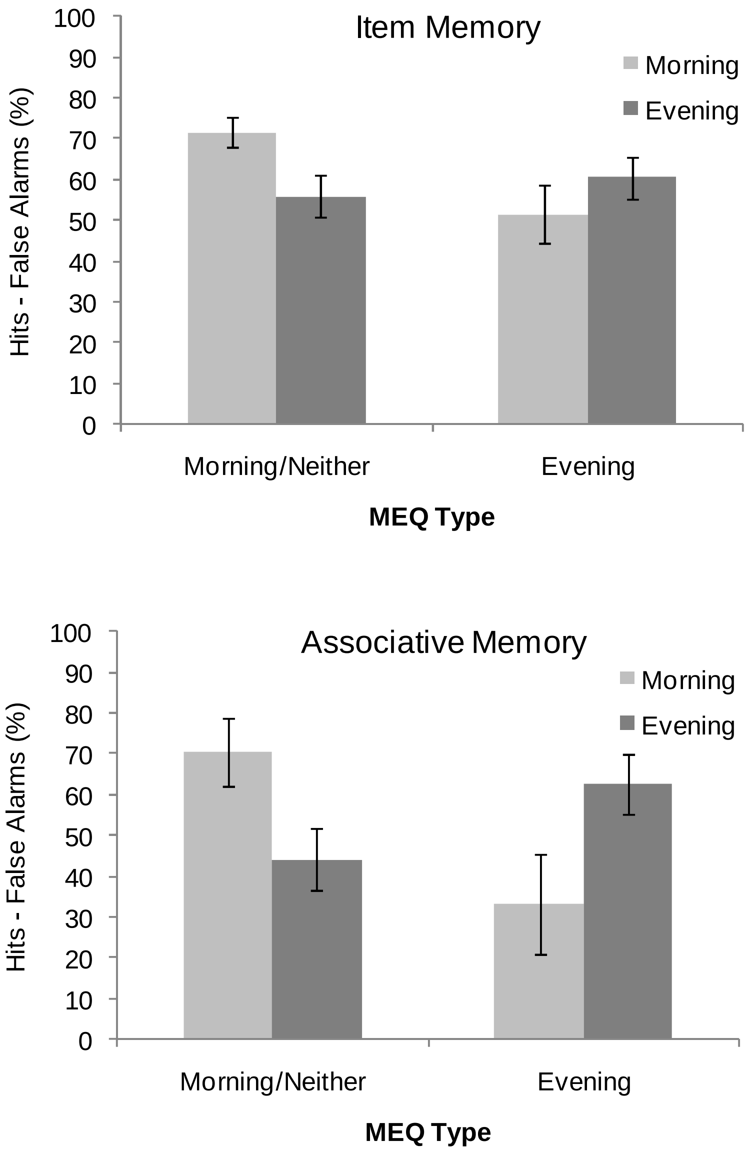

*Figure 2.* Recognition performance in Experiment 2 as percentages of hits minus false alarms for item memory (top panel) and associative memory (bottom panel) for older and young participants (morning and evening types, respectively), tested both in the morning (light bars) and in the evening (dark bars). Error bars represent ±1 *SEM*.

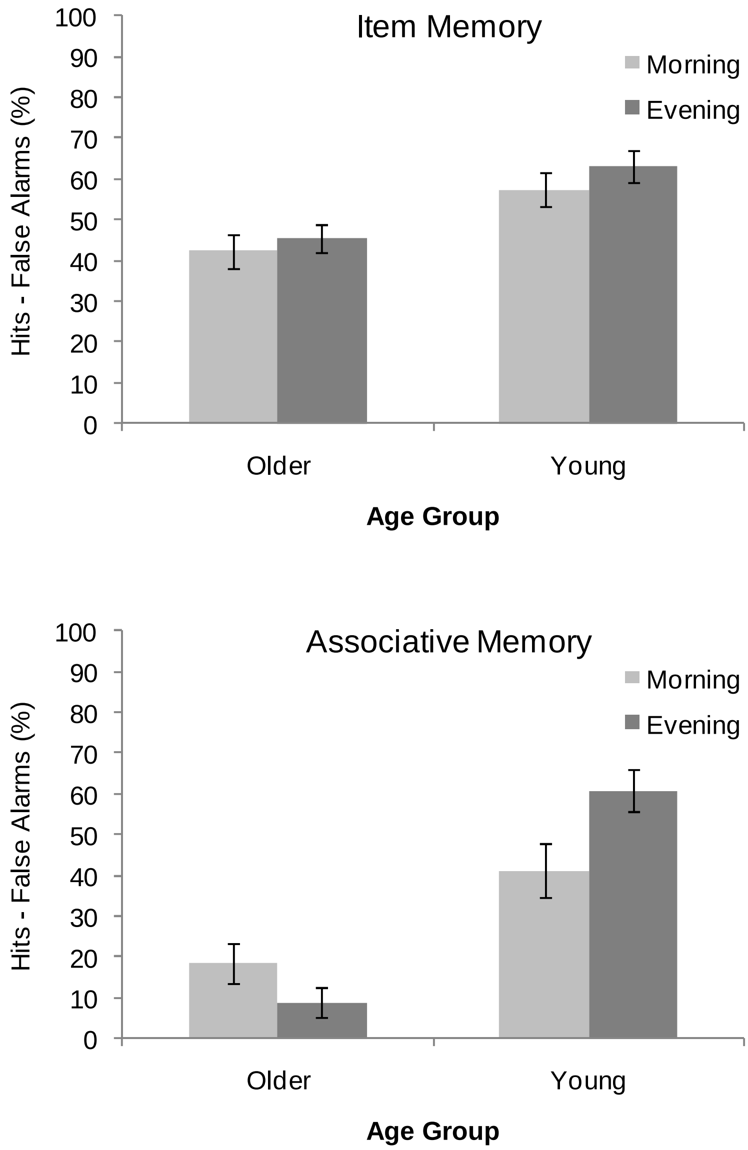

*Figure 3.* Age-related associative deficit (i.e., percentage by which the item-associative memory difference of older participants exceeded that of young participants) for each combination of young/older participants tested in the morning(AM)/evening(PM) in Experiment 2. Error bars represent ±1 *SEM*.