The experiment reported in this article examined the effects of age on memory for single words and single pictures (item information) and memory for pairs of words and pairs of pictures (associative information). Empirical differences between memory for item information and associative information have been observed at least from the 1970s (e.g., Murdock, 1974) and in more recent research (e.g., Hockley, 1991; Hockley & Cristi, 1996; Malmberg & Xu, 2007), the distinction has been brought to bear on the effects of age on cognition. Usually, only small effects of age on memory have been observed when single items are tested for recognition (Balota, Dolan, & Duchek, 2000; Craik, 1994; Kausler, 1994; Naveh-Benjamin, 2000; see the review in Neath, 1998, Chapter 16). But the differences have been significantly larger when pairs of words are tested (“Was this pair of items studied together in a list of pairs that was just studied?”; e.g., Buchler & Reder, 2007; Craik, 1983, 1986; Craik & McDowd, 1987; Healy, Light, & Chung, 2005; Naveh-Benjamin, 2000). In particular, Old and Naveh-Benjamin (2008) conducted a meta-analysis of data from 90 studies and found larger age-related deficits for associative recognition than item recognition under a wide variety of experimental manipulations.

It has usually been concluded from studies like those just cited that item information is mostly preserved with age and associative information is not. However, almost all of the studies measured only accuracy, not response times (RTs), even though older adults are typically slower than young adults. Slowing for older adults has been interpreted as a deficit such that, for example, cognitive operations are not fully completed in the available time and so the products of earlier operations are not fully available to later operations (e.g., Salthouse, 1996). When RTs are considered for item recognition, older adults do show significant deficits (Ratcliff, Thapar, & McKoon, 2004, 2010, 2011). The fact that RTs show significant deficits while accuracy does not means that a full explanation of item and associative recognition must accommodate both measures.

Tests of item and associative memory that compare older to younger adults have most often used words as the stimuli. In the experiment reported here, we also used pictures. Pictures show superior associative recognition relative to words for young adults (e.g., Hockley, 2008). If pictures allow better associative coding, then this might allow older adults to form better representations of pictures. Early studies that examined aging and item recognition with pictures (e.g., Park et al., 1986; Rybarczyk, Hart, & Harkins, 1987; Till, Bartlett, & Doyle, 1982) found only slight age-related decrements. However, Naveh-Benjamin, Hussain, Guez, and Bar-On (2003) did find an associative decrement. However, again, none of these studies measured reported instead of measured RTs. When accuracy and RTs are both considered, comparisons of item and associative memory between older and young adults face problems that require model-based analyses. Older adults’ performance relative to young adults’ is often obscured by the more conservative speed/accuracy decision criteria that they set, that is,
they are more concerned to avoid errors than young adults even if doing so slows performance (e.g., McKoon & Ratcliff, 2012, 2013; Ratcliff, Thapar, & McKoon, 2001, 2003, 2004, 2006a, 2006b, 2007, 2010, 2011; Spaniol, Madden, & Voss, 2006; Thapar, Ratcliff & McKoon, 2003; the Ratcliff et al. papers are henceforth referenced as RTM). This results in different baseline levels of performance: Older adults’ overall RTs tend to be longer and, depending on the task, their overall accuracy may be higher than young adults’ or lower. The baseline difference can lead to a scaling problem such that a larger difference between conditions in RTs or accuracy for older adults might be due to better information or it might be that they have worse information but their larger difference is due to their more conservative criteria (for further discussion, see McKoon & Ratcliff, 2012).

A direct comparison between the quality of information in memory for older and young adults requires a model that can solve the problems just described. The model must separate the quality of the information driving a decision from the effects of speed/accuracy criteria and explain how accuracy and RTs arise from the same underlying cognitive processes. We used Ratcliff’s (1978; Ratcliff & McKoon, 2008) diffusion model, a member of the class of sequential sampling models. The insights the model can offer about older adults’ memory have been demonstrated by findings that in many (although not all) memory and perceptual tasks, the quality of the information on which performance is based is as good for older as young adults (e.g., RTM papers). The reason older adults are often slower than young is largely due to the more conservative speed/accuracy criteria that they set.

The item and associative recognition tasks used for the experiment described here were two-choice tasks. Participants were given lists of pairs of items to remember. They were told whether the test would be item recognition or pair recognition only after the end of a to-be-remembered list. For item recognition, participants were asked to decide whether a test word or picture was “old” or “new” (i.e., whether it had or had not appeared in the list of pairs that was just studied). For associative recognition, participants were asked to decide whether a test pair of words or pictures was “intact” or “rearranged” (i.e., whether the two items of the test pair had or had not appeared together in the list of pairs that was just studied).

In the diffusion model (Ratcliff & McKoon, 2008) for two-choice tasks, noisy information from a stimulus is accumulated over time from a starting point until a criterion (a boundary) is reached, one criterion for each choice, at which point a response is executed. The rate at which information is accumulated is labeled “drift rate” and it is determined by the quality of the information available from the stimulus. For example, in a memory task, the quality of the information would be worse for items studied once than items studied twice. Processes outside the decision process itself are combined into one parameter of the model, “nondecision” time. The total RT for a stimulus is the time taken by the decision process plus the nondecision time. Drift rates, boundaries, and nondecision times are the three main components of the model used in understanding differences among populations of individuals and differences among individuals.

There have been four studies that have used the diffusion model to compare older to young adults’ performance on item recognition for words (Ratcliff, Thapar, & McKoon, 2004, 2010, 2011; McKoon & Ratcliff, 2012) and two that have used the model to compare performance on associative recognition (McKoon & Ratcliff, 2012; Ratcliff et al., 2011). In all cases, the older adults set their boundaries further apart and had longer nondecision times than the young adults. The important finding for understanding age differences in memory was that older adults’ drift rates for item recognition were near those of young adults, but their drift rates for associative recognition were significantly lower.

**Experiment**

There were four tasks in the experiment: item recognition with pictures and words and associative recognition with pictures and words. Our aim was to use a diffusion model analysis of the data from the four tasks to extract measures of the components of processing (i.e., parameter values) from accuracy and RT distributions for correct and error responses in order to make direct comparisons across tasks and age groups.

**Method**

**Participants**

Twenty-four college-age participants were recruited from Ohio State University and Columbus, OH, area community centers. They were paid $12 per session for their participation. Twenty-five 60–74 year olds were recruited from local senior centers and were paid $15 per session (they were paid more than the young participants because they had to travel to the location at which they were tested). The participants participated in five 55-min sessions, the first three with words as the items and the last two with pictures.

**Materials**

There were two pools of words used to make up pairs, one of 503 high-frequency words (M Kucera & Francis frequency = 217.52, range = 66–999, SD = 233.47) and one of 693 low-frequency words (M = 4.4, range = 4–5, SD = 0.5). For the picture experiments, there was a pool of 896 clip-art pictures, each a drawing of an easily recognizable object (e.g., an apple, a book, a wheelchair, a snowflake). The pictures were colored but only using 16 basic colors.

**Procedure**

For all the tasks, there was a series of study lists, each followed by a test list. A study list consisted of pairs of items and a test list consisted of either single items (“old” or “new” responses) or pairs of items (“intact” or “rearranged” responses). Participants were not informed whether items or pairs would be tested until after the study list. In addition to age and type of test item, words were presented in a study list once or twice, pictures were presented once or twice, and words occurred with low frequency or high frequency in English. These variables were added to provide more constraints on the diffusion model’s ability to fit the data.

**Results**

The data for the three sessions with words were combined for analyses, as were the data for the two sessions with pictures. RT cutoffs of 400 ms and 3000 ms were used for the young adults and
and materials, respectively. All of the interactions were also significant: age and task, \( F(1, 47) = 44.1, \eta^2 = .04 \). As noted in the introduction, pictures might have supported memory for associations for the older participants such that the decline in accuracy with age for associative recognition would be less for pictures than words, but this was not the case. There was a significant effect of material type on performance, \( F(1, 47) = 97.8, \eta^2 = .19 \), and a significant effect of task on performance, \( F(1, 47) = 84.5, \eta^2 = .84 \), but these are not interesting because it is difficult to equate pictures and words and tasks. There were also interactions of task by material type, \( F(1, 47) = 10.5, \eta^2 = .004 \), and a triple interaction of age, task, and material type, \( F(1, 47) = 9.5, \eta^2 = .004 \), but both of these were small.

For RTs, the main effects of age, materials, and task were significant—\( F(1, 47) = 86.9, \eta^2 = .37 \), \( F(1, 47) = 82.8, \eta^2 = .09 \), and \( F(1, 47) = 6.8, \eta^2 = .003 \), respectively—showing that older adults were slower than young adults, and mean RTs differed for the tasks and material types (see Figure 1). The slow responses for pairs of pictures compared to pairs of words may be partly due to the difference in testing procedure: For words, the first word of a test pair was presented in advance of the second word whereas for pictures, the two pictures of a test pair were presented simultaneously. All of the interactions were also significant: age and materials, \( F(1, 47) = 114.4, \eta^2 = .12 \), showing that response times were different for pictures and words as a function of age; age and task, \( F(1, 47) = 63.6, \eta^2 = .03 \), showing that RTs differed for item and associative recognition with age; task and materials, \( F(1, 47) = 301.5, \eta^2 = .09 \), showing that RTs for pictures and words differed as a function of task type; and age, task, and materials, \( F(1, 47) = 21.7, \eta^2 = .007 \).

**Diffusion Model Analyses**

For mean drift rates (averaged over conditions), boundary settings, and nondecision times, Figure 2 shows the best-fitting values averaged over participants for pictures and words and for the older and young participants. Goodness of fit is discussed in detail in the supplement. For the analysis of variance results below, we report values that are significant at the .05 level.

There was a marginally significant effect of age on drift rates, \( F(1, 47) = 3.4, p = .07, \eta^2 = .02 \). But there was an interaction between age and task with a greater decline in drift rates for associative recognition than item recognition, \( F(1, 47) = 7.9; p < .05, \eta^2 = .01 \), showing the older adults’ difficulty with associative memory. Drift rates were larger for item recognition than associative recognition, \( F(1, 47) = 202.4, \eta^2 = .19 \), and drift rates were larger for pictures than words, \( F(1, 47) = 87.9, \eta^2 = .24 \). The Material Type × Age interaction was not significant, \( F(1, 47) = 0.4 \), showing that drift rates were not different for young and older adults as a function of the material type. The Task × Material Type interaction showed a larger difference between pictures and...
words for single-item tests than pairs tests, \( F(1, 47) = 94.7, \eta^2 = .07 \). This interaction is not particularly interesting because it is a function of the study time and item type and it is difficult to equate pictures and words. The triple interaction between age, material type, and task was not significant, \( F(1, 47) = 2.1 \).

For boundary settings, the effects of age, materials, and tasks on boundary placement were straightforward: boundaries were set further apart by older than young participants, \( F(1, 47) = 25.2, \eta^2 = .26 \), they were set further apart for words than pictures, \( F(1, 47) = 80.9, \eta^2 = .08 \), and they were set further apart for pairs tests than single-item tests, \( F(1, 47) = 17.6, \eta^2 = .02 \). None of the interactions among these variables was significant, \( F(1, 47) < 1.5 \).

As is usually found, nondecision processes took more time for the older than the young participants, \( F(1, 47) = 39.5, \eta^2 = .28 \). There were main effects for material type, \( F(1, 47) = 36.4, \eta^2 = .05 \), and task, \( F(1, 47) = 36.7, \eta^2 = .07 \). These main effects were qualified by an interaction, \( F(1, 47) = 141.1, \eta^2 = .09 \), in which nondecision times for pictures were about the same as for words with single-item tests, but they were longer for pictures than words for pairs tests. As noted earlier, this result is likely due to the method of presentation: for words, one member of the pair was presented in advance of the second, whereas for pictures, both members were presented at the same time. There was also a marginally significant interaction between age and task, \( F(1, 47) = 3.4, p = .07, \eta^2 = .01 \) (Figure 2). The interactions between age and materials and among age, task, and materials were not significant, \( F(1, 47) < 1.8 \).

**Differences Among Individuals in Diffusion Model Parameters**

Even though there were relatively few participants in these tasks for individual difference analyses, there were strong effects that replicated across the different combinations of tasks. Such individual differences are important in understanding factors that go beyond a single experiment and a single performance measure. Here we summarize results that are presented in full in the supplement.

First, we would expect that faster participants would be more accurate, but there is no correlation between accuracy and RT across participants for any of the tasks (the mean is \(-.04\)). The diffusion model explains this because boundary settings, nondecision time, and drift rates are not correlated, drift rates largely determine individual differences in accuracy, and boundary separation and nondecision time largely determine individual differences in RT. Second, there were strong correlations across tasks for both participant groups in drift rates, as well as in boundary separation and in nondecision time. This means that if a participant had a high value in one task, they had a high value in the other tasks. Third, there were reasonably strong correlations of drift rates with matrix reasoning IQ measure and of drift rates with verbal IQ, except for young adults on the picture task drift rates.

**General Discussion**

The differential decline in accuracy for item versus associative memory as a function of age has been an important finding for hypotheses about how memory changes with age, but such hypotheses have not usually been targeted at differences in RTs. The

---

*Figure 2. Plots of drift rate, boundary separation, and nondecision time for item recognition and associative recognition for picture and word stimuli for each of the participant groups.*
data from our experiment showed age effects on both and the diffusion model analyses of the two measures, the two tasks, and the two types of items allow them to be understood in terms of components of processes that underlie decisions. A key question was whether RT data change how the effects of aging on cognition are understood.

A crucial feature of the diffusion model is that the information from memory that determines decisions does not map directly to responses. Instead, the decision process intervenes; it transforms the information from memory to a response. Whether a particular response is chosen depends not only on the amount of information in favor of it but also on the boundaries that set the amount of information that must be accumulated before a response can be executed.

For the experiment reported here, for each task and participant, the diffusion model was fit to all the data simultaneously (accuracy and RT distributions for all the conditions), and the predictions from it matched the data well (see the supplement). As is usual, the data provide tight constraints on the model. The strongest are that it must produce the right-skewed shapes of RT distributions and the appropriate effects of independent variables on them. Specifically, as the difficulty of test items increases, increases in mean RTs must come from increases in the spreads of the distributions, not changes in their leading edges. Ratcliff (2002) has shown that, although there are other plausible ways that the distributions could change with difficulty, they are not observed in real data.

The differences in the patterns of accuracy and RTs were the same for pictures as for words, and they are easily summarized. For item recognition, the older adults’ accuracy was about the same as the young adults, but their RTs were longer. For associative recognition, the older adults were both less accurate and slower than the young adults. The increases in RTs from young to old were large, ranging from 200 ms to 400 ms.

In terms of underlying components of processing, older adults were slower than young adults in both item and associative recognition because they set their boundaries further apart and their nondecision times were longer. For item recognition, they were as accurate as the young adults mainly because their drift rates were about the same, and for associative recognition they were less accurate mainly because their drift rates were lower.

Starns and Ratcliff (2010, see also 2012) explored what aspects of performance older and young participants in their experiments might be trying to optimize with their boundary settings. They defined optimality in terms of “reward rate” (the term comes from animal research; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006), where reward rate is the number of responses that are correct per unit time. They found that young participants set their boundaries at values that give close to optimal performance when accuracy feedback is given on each trial and they have had extensive practice. In contrast, older participants set their boundaries such that they obtain close to the maximum possible accuracy; they do not adjust their boundary settings even with accuracy feedback and extensive practice. The interpretation of this is that older adults were unwilling to make errors that they can avoid, that is, responses for which they might realize a few moments later that they were incorrect. This difference in optimality goals might be part of a general explanation of why older participants are slower than young adults.

The results of our experiment showed associative deficits for the older relative to the young adults that were similar for pictures and words. If it is easier to encode pictures than words, then it might be thought that pictures would be easier to bind into associative relations than words, but our data did not show such an effect. In contrast, Naveh-Benjamin et al. (2003) and McKoon and Ratcliff (2012) found benefits when the pairs of words to be learned were related to each other (e.g., “dog” and “cat”) than when they were not.

Summary

In any task, laboratory or real world, decisions take time. Sequential sampling models like the diffusion model integrate RTs and accuracy into a single framework that specifies how the two dependent variables come from the same underlying components of decision making. Because the diffusion model accounts for accuracy and RTs simultaneously, it allows researchers to look beneath surface data, that is, beneath accuracy and RTs, to the information from memory on which decisions are based. Perhaps the most important implication is that when an explicit decision-making model like the diffusion model is applied to accuracy and RT data, simple interpretations based on one or the other of the variables can be replaced by coherent interpretations based on all the data.

Our results show that the diffusion model fits data from item and associative recognition with both picture and word stimuli. The fits of the model show that evidence used in making the decision for item recognition is affected only modestly with age, whereas evidence for associative recognition declines significantly with age both for picture stimuli and word stimuli. RT differences between older and young adults are explained by older adults adopting more conservative decision criteria and having longer nondecision times. These model components (drift rates, criterion settings, and nondecision times) are strongly correlated across tasks, which means that across individuals, there are consistent individual differences. However, the model components within each task are not correlated with each other, and this provides an understanding of why accuracy and RT are not correlated across participants.

References


