Received Date : 21-Mar-2016
Revised Date : 30-May-2016
Accepted Date : 31-May-2016
Article type : Paper (under 8000 words)

Comparing single- and dual-process models of memory development

Brett K. Hayes[1], John C. Dunn[2], Amy Joubert[1], and Robert Taylor[1]

[1] University of New South Wales, Australia
[2] The University of Adelaide

Author Note

This work was supported by Australian Research Council Discovery Grant DP150101094 to the first and second authors. We thank Jeremy Ngo for help with programming the experiments and in manuscript preparation. We are very grateful to the students and staff of Oxley College Primary School, Bowral, and Burgmann Anglican School, Canberra for their enthusiastic participation.

Correspondence concerning this article should be addressed to Brett K. Hayes, School of Psychology, The University of New South Wales, Sydney, NSW, 2052, Australia. E-mail: B.Hayes@unsw.edu.au

Research Highlights

Single- and dual-process models of memory development

- Comprehensive comparison of single-process and dual-process models of memory development from 6-years of age to adulthood
- Uses state-trace analysis to identify the minimum number of latent variables needed to explain memory development
- Results show that single-process memory models based on memory strength provide a better account of memory development than dual-process models

Abstract

This experiment examined single-process and dual-process accounts of the development of visual recognition memory. 6-7-year-olds, 9-10-year-olds and adults were presented with a list of pictures which they encoded under shallow or deep conditions. They then made recognition and confidence judgments about a list containing old and new items. We replicated the main trends reported by Ghetti and Angelini (2008) in that recognition hit rates increased from 6- to 9-years of age, with larger age changes following deep than shallow encoding. Formal versions of the dual-process high threshold signal detection model and several single-process models (equal variance signal detection, unequal variance signal detection, mixture signal detection) were fit to the developmental data. The unequal variance and mixture signal detection models gave a better account of the data than either of the other models. A state-trace analysis found evidence for only one underlying memory process across the age range tested. These results suggest that single-process memory models based on memory strength are a viable alternative to dual-process models for explaining memory development.

Comparing single- and dual-process models of memory development

Two enduring issues in the study of memory development are the number of qualitatively distinct processes needed to explain how children recognize items and the developmental trajectory of these processes. Single-process models assume that recognition decisions are based on an assessment of the strength of a memory signal elicited by a test item (Dunn, 2008; Wixted, 2007). If this strength exceeds some subjective response threshold the item will be recognized. Such models assume that the memory strength of studied items relative to those that have not been studied increases with age.

Dual-process accounts on the other hand, assume that recognition is based on two qualitatively different processes. Like single-process models there is an assessment of the memory strength or *familiarity* of an item. Additionally however, there is a process of *recollection*, involving conscious retrieval of the episodic details associated with the presentation of a study item (Jacoby, 1991; Yonelinas, 1994, 2002). Dual-process theories of memory development assume that recollection emerges later and shows more marked developmental change than familiarity (Brainerd, Holliday & Reyna, 2004; Ghetti & Angelini, 2008; Koenig, Wimmer, & Hollins, 2015). Specifically, the development of familiarity is thought to stabilize around 7 or 8 years of age whereas recollection continues to improve through childhood and adolescence (Ghetti & Lee, 2011).

In the adult memory literature there is still heated debate about which approach provides the best account of memory (cf. Pazzaglia, Dube & Rotello, 2013; Yonelinas & Jacoby, 2012). Nevertheless, dual-process accounts of memory development have gained considerable popularity. For example, a review of research on episodic memory development confidently concluded that “age-related improvements in memory performance are driven by the development of recollection” (Ghetti & Lee, 2011, p. 266). Such conclusions are based on three forms of evidence; i) behavioral studies which claim to show developmental change in recollection from the early school years to adolescence and adulthood together with relative stability in familiarity; ii) computational modelling which shows that these developmental data can be fit by a formal dual-process model, and iii) studies which show age related changes in the neural correlates of recollection.

The aim of the current work was to re-examine the first two types of evidence for dual-process models of memory development, and to offer an alternative single-process account. First we replicated the key behavioral findings of an important study (Ghetti & Angelini, 2008) that is seen as supporting the dual-process account. Second, in contrast to previous work, we fit a variety of psychologically plausible single-process models, as well as a popular dual-process model, to the recognition data from children and adults. To pre-empt our main finding, variants of the single-process approach provided a better account of the developmental data than the dual-process model. The current work does not examine the neural correlates of memory but the implications of our findings for such work are addressed in the Discussion.

Behavioral and modelling evidence supporting dual-process accounts of memory development

Much of the evidence for dual-processes in children’s memory has been obtained using adaptations of paradigms used to study adult memory. One popular approach is the remember-know paradigm in which participants are asked to report whether or not a recognized test item is accompanied by memory of the specific details associated with the study episode (Gardiner & Richardson-Klavehn, 2000). If this occurs, then the participants are said to “remember” the item, and are assumed to be relying primarily on recollection. If not, then participants are said to “know” that they have seen the item before, and are assumed to be relying on familiarity. Developmental studies have found different age patterns for remember and know judgments. Billingsley, Smith, and McAndrews (2002), for example, found that the proportion of correct “remember” responses to previously presented pictures increased steadily from 8 to 19 years of age. In contrast, the proportion of “know” responses remained constant across this period (see Piolino, Hisland, Ruffeveille, Matuszewski, Jambaqué, & Eustache, 2010 for similar results). Ghetti, Mirandola, Angelini, Cornoldi and Ciaramelli (2011) developed a modified version of the remember/know instructions (relabeled as “remember” vs. “familiar”) that could be understood by children as young as six years of age. Remember responses to correctly recognized items followed a U-shaped age pattern, increasing from 6 to 9 years but decreasing in adulthood. Familiar responses were relatively rare (e.g., constituted only 12% of 6-year-olds’ responses) and were stable across childhood.

However using the remember-know procedure to study the processes that underlie memory development is problematic. The metacognitive ability to introspect about memory states such as “remember” versus “know”, has been shown to undergo marked developmental change from 6 years to adulthood (Ghetti, Qin & Goodman, 2002; Ghetti, et al., 2011; Perner & Ruffman, 1995). It is difficult to disentangle such metacognitive changes from genuine changes in the processes that underlie recognition. More generally, this approach relies on the logic of empirical dissociations to draw conclusions about the processes that drive recognition. Different processes are inferred when variables like age are shown to affect one type of memory judgment (e.g., remember responses) but not others. Such logic however has been shown to be flawed, with numerous demonstrations in the adult memory literature that such dissociations can arise from a single underlying process (e.g., Dunn, 2008; Pratte & Rouder, 2012; Prince, Brown, & Heathcote, 2012).

An alternative approach that avoids many of these problems was used by Ghetti and Angelini (2008). Children aged 6, 8, 10, 14 and 18 years were presented with 160 line pictures under either shallow (e.g., “what color ink is this drawing?”) or deep encoding (e.g., “was the thing in the drawing light or heavy?”) conditions. At test participants made recognition judgments about 160 studied target items and 80 unpresented lures, and rated their confidence in these judgments. Hit rates (correct recognition of targets) increased with both age and with deep encoding. These factors also interacted, with larger age increases in hits following deep as compared with shallow encoding.

Notably, recognition judgments and confidence ratings were used to generate receiver-operating-characteristic (ROC) curves for participants in each condition. ROC curves are plots of recognition hit rate against false alarm rates at different levels of decision confidence. A formal version of a dual-process model, the high-threshold signal detection model (HTSD) proposed by Yonelinas (1994), was then fit to the ROC data. The HTSD model holds that recollection either fails or succeeds for a given test item. When recollection fails, recognition is based on a continuous familiarity process governed by an equal-variance signal-detection model. When recollection succeeds, an "old" decision is made with high confidence and high accuracy. Recollection and familiarity are assumed to contribute independently to recognition (see Appendix A for model equations). The key model parameters correspond to familiarity (d) and the probability that some information from the study episode is recollected (r).

The HTSD provided a good overall fit to the developmental recognition data. Moreover the pattern of model parameters across conditions was in line with the assumptions of a dual-process model. The recollection parameter of HTSD increased from 6-8-years to 10-years of age in the deep but not the shallow encoding condition. The familiarity parameter was also larger following deep as compared with shallow encoding, but showed less developmental change than recollection. Six year olds showed lower levels of familiarity than older children.

Single-process models of memory development

The results of the model fitting carried out by Ghetti and Angelini (2008) are consistent with a dual-process account of memory development. Crucially, however because the fit of the HTSD model was not compared with model fits based on formal single-process models, the issue of the number of processes underlying the development of recognition remains open. What is needed to advance our understanding of this issue is a *comparison* of the relative fit of single-process and dual-process models to developmental data on recognition decisions and confidence judgments. The current study therefore sought to a) replicate the key developmental trends in the Ghetti and Angelini (2008) study[1], b) compare the quantitative fits to these data of a dual-process model (HTSD) and a number of single-process signal detection models, and c) apply state-trace analysis (Dunn, 2008) to determine if a dual-process model is *necessary* to account for the findings.

Single- and dual-process models of memory development

Work on adult memory suggests that single-process models vary in their ability to account for recognition data. Early signal detection models of memory assumed that recognition decisions were made with reference to two equal-variance Gaussian distributions, one representing presented previously target items and the other representing unpresented lures (see Figure 1, top panel). A body of evidence however suggests that such equal-variance signal detection models (EVSD) often do not provide a good account of adult recognition (e.g., Ratcliff, Sheu, & Gronlund, 1992; Wixted, 2007). For example, the EVSD predicts that adults' ROC curves will have a symmetric shape whereas most studies have found curvilinear ROCs.

An alternative single-process approach, the unequal-variance signal detection model (UVSD), makes additional assumptions about the relative variance of the distribution of memory strengths for targets and lures. The presentation of items during a study period should add memory strength to those items relative to lures. However it seems reasonable to expect that not all studied items will show the same increase; some will benefit more from study presentation than others. Hence, the UVSD model assumes that the variance of memory strengths for targets will be higher than that of lures (see Figure 1, bottom panel). The formal model (see Appendix) is structured around two key parameters, the mean distance between the distributions of memory strength for targets and lures (d') and the variance of the target distribution (s).

A related single-process model, the Mixture signal detection model (MixSD, Decarlo, 2002) also assumes that memory strength is the key determinant of recognition memory. In this model participants either do or do not pay attention to targets during a study phase. If they do the target receives a constant increase in memory strength. However if they don't pay attention there is no increase in strength. The resulting distribution of responses for target items therefore is a mixture of the distributions for targets and lures from the equal variance model. Formally, the model has two key parameters, memory strength parameter d' and the probability of attending to the study item λ (see Appendix A for more details).

Both the UVSD and the MixSD models predict curvilinear ROCs and have generally outperformed the EVSD in capturing adult recognition patterns (DeCarlo, 2002; Wixted, 2007). Notably, in many cases these models have been shown to outperform the dual-process HTSD model in fits to adult recognition (e.g., Rotello, MacMillan, Reeder & Wong, 2005; Wixted, 2007).

In the current study we compared the quantitative fits of three single-process models (EVSD, UVSD, MixSD) and the HTSD dual-process model to visual recognition data from adults and children. These data were obtained using a visual recognition memory task similar to that used by Ghetti and Angelini (2008). The two groups of children tested (6-7 year-olds and 9-10 year-olds) were those who showed the largest age-related changes in hit rates in Ghetti and Angelini (2008). Adults were also included as a comparison group. As per Ghetti and Angelini (2008), participants were presented with a list of pictures and given either shallow or deep encoding instructions on different study blocks. They then made recognition judgments and gave confidence ratings to a test set containing target and lure pictures. Based on Ghetti and Angelini it was expected that 1) hit rates would increase with age, 2) hit rates would be higher following deep compared with shallow encoding, and 3) the effect of encoding depth would be greater for older children and adults than for younger children.

In addition to comparing overall model fits we also examined how the key parameters of each model changed with age and depth of encoding. According to dual-process accounts age increases in hit rates following deep encoding will be reflected in increases in the recollection (*r*) parameter. In contrast single-process signal-detection models would generally predict that increases in hit rates due to age and depth of encoding will be reflected by substantial changes in *d*'.

To further examine the number of processes that underlie memory development we applied a state-trace analysis to child and adult recognition data. State-trace analysis (Bamber, 1979; Dunn, 2008; Prince et al., 2012) is based on the premise that two dependent variables will covary to the extent that changes on the two variables are mediated by the same latent variable. By producing a plot of one dependent variable (e.g., high confidence hits in recognition memory) as a function of another dependent variable (e.g., low confidence hits in recognition memory), one can determine the number of latent variables (or psychological processes) needed to account for the observed data. A monotonic plot suggests that the data are mediated by a single underlying process. However, if the plot shows non-monotonic discontinuities, more than one underlying process is involved.

This approach has been used to examine rival single- and dual-process claims concerning adult memory (e.g., Dunn, 2008), face processing (e.g., Prince et al., 2012) and categorization (e.g., Newell, Dunn & Kalish, 2011). STA has proven to be a useful tool in diagnosing the underlying dimensionality of such tasks, often revealing that data which has been seen as supporting dual-process accounts (such as dissociations between remember and know recognition judgments) can be generated from a single underlying process (Dunn, 2008). To our knowledge, this is the first time STA has been applied to memory development data.

Method

Participants

Ninety-six participants from three age groups were recruited. Thirty-two 6-7 year olds (M_{AGE} = 6 years, 9 months; Range = 73 months - 89 months; 19 females) and thirty-two 9-10 year olds (M_{AGE} = 9 years, 5 months; Range = 106 months -122 months; 17 females) were recruited from two private primary schools in a rural area. All participants were from middleclass families. Thirty-two undergraduate psychology students (M_{AGE} = 19 years, 1 month; 19 females) participated for course credit. No participants were excluded from data analyses.

Materials

Pictorial stimuli. One hundred and twenty-eight line pictures, previously normed for visual complexity, familiarity and name agreement with child participants (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997) were used. Four lists of 32 pictures were created by allocating pictures to the four categories used in the deep encoding study blocks (living, non-living, indoor items, outdoor items). Half of the pictures in each list were randomly assigned as study items and half as test items. Half of the study pictures were further randomly assigned to either shallow or deep encoding blocks. Hence there were four versions of the study/test items, which were counterbalanced across participants in each age group. Half of the pictures shown during the study phase were in green ink and half in red ink.[2]

Procedure

The procedure was patterned after Ghetti and Angelini (2008, Experiment 1) with differences as noted. Children were tested individually in a quiet room at their school and were guided through all stages of the experiment by the second author. Adults were tested individually in laboratory cubicles. All stimuli and instructions were presented on a computer screen and responses were made using the computer mouse. The experimenter read instructions aloud for child participants and recorded responses on their behalf.

The experiment took place in four phases; practice, study phase; confidence-scale training and recognition test. The practice phase used pictures from the Cycowicz, et al., (1997) set that were not used in the rest of the experiment. Participants were told that they would see pictures that were in either green or red ink. For each picture, they were told they would be asked to either (a) name the color it was drawn in (shallow encoding) or (b) answer a question based on its color (deep encoding). For pictures in green ink the participant had to judge whether the object was living or non-living. For pictures in red ink the participant had to judge whether the object was generally found indoors or outdoors. During encoding practice, each participant completed 6 training trials; 2 shallow encoding and 4 deep encoding (2 trials for each of the two types of semantic judgments).

In the study phase participants were shown 64 pictures. Items were presented in shallow or deep encoding blocks of 32 items, with block order counterbalanced across participants. The presentation order of pictures within each block was randomized. Participants did not receive a break between shallow and deep encoding. Each picture was presented in the center of the computer screen for 1.5 s, followed by a white screen. The level of encoding question (shallow: “was the thing in the drawing red or green?” deep: “was the thing in the drawing alive or not alive?” or “was the thing in the drawing usually found indoors or outdoors?”) was presented after each item. The experimenter also delivered this encoding question verbally for child participants. Participants did not receive feedback on their responses. After the encoding response, there was a 0.5 s interval during which the screen remained blank and then the next item was shown.

Immediately after the study phase, participants received training on a confidence rating scale (adapted from Berch & Evans, 1973 and Ghetti, et al., 2002). Children were instructed to point to a photo of a confident-looking child if they felt sure about their answers. They were told to point to a photo of an uncertain-looking child if they felt unsure about their answer. The gender of the photos was matched to that of each participant. Adults were shown two buttons, labeled “sure” and “unsure”, and instructed to click the one that best represented their level of confidence about their recognition judgment. Child participants received four confidence practice trials. On each trial a fragmented picture from the Cycowicz, et al., (1997) set not used in other phases was presented. Participants were asked “Is this an [object name]?” and to choose the appropriate level of confidence for their judgment. On two trials only a small proportion of picture pixels were deleted, meaning the picture was easy to identify. On two trials a large proportion of pixels were deleted and the picture was hard to identify. These trials were respectively used to illustrate “sure” and “unsure” levels of confidence.[3]

The recognition test phase followed with the presentation of 128 pictures in black ink. Item presentation order was randomized and progress through the recognition test was self-paced. For each picture, participants had to judge whether the picture was “old” or “new” and give a binary confidence rating for the judgment. Children were required to point to the appropriate photograph on the confidence scale whereas adults responded by clicking “sure” or “unsure” buttons. If a picture was judged to be “old”, participants were also asked whether they could remember if the picture was originally shown in red or green ink. Accuracy on this task was used as an additional measure of memory for episodic details (Ghetti & Angelini, 2008).

Overall our study/test set differed from Ghetti and Angelini (2008) in two ways. First, we used fewer study and test items (total of 128 as compared with 240). We did this because Ghetti and Angelini found that, a) hit rates declined in the second half of their recognition test, presumably due to fatigue, and b) their key age and encoding effects were still observed in the first half of their test. The length of our test set was therefore similar to the first half of their test set. Second, in our study an equal number of target and lure items were presented at test, compared to Ghetti and Angelini (2008) who used a ratio of 2:1 for targets and lures. We took this step to minimize the chance that our results would be contaminated by a positive response bias.

Results

Recognition Judgments

Preliminary analyses confirmed that the proportion of target items correctly identified as old (hits), and lures incorrectly identified as old (false alarms), were unaffected by participant gender (F's <1.0). They also confirmed that recognition rates were unaffected by the version of study/test items used or the order of encoding blocks during study (F's <2.5). All subsequent analyses were collapsed over these factors.

Hit and false alarm rates are given in Table 1. Hits were entered into a 3 (age) x (2) (encoding) analysis of variance (ANOVA) with repeated measures on the second factor. Planned comparisons were used to examine age differences between 6-7-year-olds and 9-10-year-olds, and between 9-10-year-olds and adults respectively. There was a large effect of encoding level with hit rates higher following deep (M = 0.78) than shallow encoding (M = 0.52), $F(1, 93) = 287.79$, $p<.001$, $\eta_p^2 = 0.76$. Hit rates increased significantly from 6-7 (M = 0.58) to 9-10-years of age (M = 0.68), $F(1, 93) = 10.94$, p=.001, $\eta_p^2 = 0.11$, but did not differ between 9-10 years and adults (M = 0.69), $F(1, 93) = 1.55$, p = .22. However, there was a significant interaction between level of encoding and the age increase from 6-7 to 9-10 years, $F(1, 93) = 4.16$, p=.048, $\eta_p^2 = 0.04$. The table shows that the age increase in hit rates was larger following deep than shallow encoding. The table also shows that false alarm rates were relatively low. False alarms were unaffected by level of encoding or age (F's <1.0).

These hit and false alarm rates are similar to those of the closest corresponding age groups in Ghetti and Angelini (2008) and replicate their main qualitative findings. Recognition performance was better following deep than shallow encoding and increased over the early elementary school years. Moreover these factors interacted with a larger age increase following deep than shallow encoding.

Confidence Ratings

Before proceeding to the generation of ROC curves based on recognition and confidence responses it was important to check that the three age groups were using the confidence scale in a similar way. The proportion of "sure" responses was computed separately for hits and false alarms. As shown in Table 2, confidence for hits was higher following deep than shallow encoding, $F(1, 93) = 303.12$, $p<.001$, $\eta_p^2 = 0.77$. There were no effects of age on confidence for hits and no age by encoding interactions (all F's <2.4). The table shows that participants rarely made a "sure" response to false alarms and that this did not vary with age ($F< 1.1$).

Recall of Item Color

Memory for the color of old items provides a further measure of age and encoding effects on retrieval of episodic features. Memory for item color was calculated as the number of times that a participant correctly identified item colors divided by the number of old items they correctly reported as old, so that accuracy ranged from zero to one. Color memory was above chance in each age and depth condition (Table 3). The majority of accurate color identifications (82%) were made when participants were "sure" of their old response. As shown in Table 3, color memory accuracy increased with age, $F(2, 93) = 7.18$, $p=.001$, $\eta_p^2 = 0.13$, and with deep encoding, $F(1, 93) = 10.21$, $p<.002$, $\eta_p^2 = 0.31$. However these effects were modified by interactions between encoding and the comparison between 6-7-year-olds and 9-10-year-olds, $F(1, 93) = 8.71$, $p=.004$, $\eta_p^2 = 0.09$, and between encoding and the comparison between 9-10-years-olds and adults, $F(1, 93) = 12.47$, $p=.001$, $\eta_p^2 = 0.11$. The table shows that there was little age change in accuracy following shallow encoding, whereas accuracy increased linearly with age following deep encoding. These results also parallel those of Ghetti and Angelini (2008), who found more marked age improvements in memory for color following deep than shallow encoding.

ROC Curves and Model Fitting

ROC curves were generated by plotting the cumulative proportion of hits and false alarms for each level of confidence in each age and encoding condition (see Figure 2). Four models (equal variance signal detection, unequal variance signal detection, mixture signal detection model and high threshold signal detection) were fit to the ROC responses to targets under shallow and deep encoding, and responses to lures. Simultaneous fitting of deep and shallow targets meant that for each age group there were six categories of responses (old shallow-sure, old shallow-unsure, old deep-sure, old deep -unsure, new-sure, new-unsure). For all models a common set of decision criteria (*c1, c2, c3*) were estimated for shallow and deep conditions. This simplification is justified given previous evidence showing that people rarely shift their recognition decision criteria in response to within-list manipulations (e.g., Stretch & Wixted, 1998). All other model parameters were estimated separately for the shallow and deep conditions.

Model fitting was carried out separately for each age group, using the R package (R Core Team, 2013). Modelling proceeded by an iterative search for the free parameter values that maximized the log likelihood of the fit of each model (Myung, 2003). Goodness of fit was indexed using the G^2 statistic, which measures the fit of the model based on maximized log likelihoods relative to an optimal model that fits the data perfectly. A smaller G^2 indicates better fit. For large samples G^2 is distributed as a chi-square statistic so that a significance test can be applied for each model fit. The estimated parameter values and goodness of fit measures for each model are shown in Table 4.[4]

It is well accepted that goodness of fit statistics alone provide at best, only part of the answer to how well a model explains a given data set (Pitt, Myung & Zhang, 2002). Arguably a more important index is the *relative* fit of each model, which was evaluated by calculating the Akaike information criterion (AIC, Akaike, 1973) and Bayesian information criterion (BIC, Schwartz, 1978) for each. These measures assess relative model fit, after adjustment for the number of free parameters. The latter feature is important because we were comparing models with different numbers of parameters (the EVSD model has 5 free parameters; UVSD, MixSD and the HTSD model each have 7 free parameters). The AIC and BIC results are given in Table 5, with lower values indicating better model fit.

Table 4 shows that the EVSD model provided a poor fit to the recognition data, with significant discrepancies (as measured by G^2) between the model predictions and the data for each age group. Notably, the dual-process model also failed to provide an adequate fit to the data from any of the three age groups. The two more complex single-process models fared somewhat better. For all age groups UVSD model and MixSD models produced a better approximation of the data, as measured by as measured by G^2, than either the EVSD or HTSD dual-process models. Both of the more complex single-process models provided an overall fit to the recognition data for 6-7-year-olds that did not differ from an ideal model. For the MixSD this was also the case for the adult recognition data.

It was surprising that no model provided an accurate fit of the 9-10-year-old data. This was largely due to all models underestimating 9-10-year-olds’ proportion of “old unsure” responses to target items, especially following deep encoding. Inspection of the data suggested that this pattern may have been influenced by an outlier in the 9-year-old group who made far more "old unsure" responses to deeply encoded targets (59% of target responses) than any other participant in any age group (mean "old unsure" target responses across ages = 11%). When this outlier was removed, the fit of all four models to the 9-year-old data improved (see Appendix B for details). Notably in this case, both the MixSD and UVSD models provided an accurate fit (i.e. their predictions did not differ significantly from an ideal model).

In terms of relative model fit (Table 5), both the AIC and BIC indexes showed that the UVSD and MixSD models provided a better account of the data than the dual-process model or EVSD. In terms of the criteria for evaluating AIC differences suggested by Burnham and Anderson (2002), there was "considerably less support" for the dual process model than the mixture model in fits to the 6-7-year-old data and "essentially no support" for the dual process model over the mixture model in fits to 9-10-year-old and adult data. There were relatively small differences in the relative fit of the mixture and UVSD models, but these consistently favored the mixture model.

We now turn to an examination of the model parameters estimated in each age and encoding condition. The dual-process model predicts that the contribution of recollection to recognition should increase with age and deep encoding. Table 4 shows that trends for the dual-process r parameter did indeed increase with age and encoding depth. Notably however, there was an also an increase in familiarity (as measured by d') with deeper encoding and an increase in familiarity from 6-7 to 9-10 years of age. As detailed in the next section, this trend is potentially problematic for the dual-process model because the age increase in both parameters can be explained parsimoniously by models with a single underlying process.

In both the UVSD and mixture models, d' increased with age, with the largest increases observed between 6-7 and 9-10 years. d' was also higher following deep as compared with shallow encoding. In UVSD, the s parameter remained relatively stable across age and depth of encoding conditions. In the mixture model the λ parameter was relatively stable across the three age groups. In all three groups λ was higher following deep than shallow encoding. Hence, in both UVSD and mixture models, age changes in recognition memory were largely the result of changes in memory strength. In the mixture model there was also evidence that deep encoding increased attention to study items in each age group.

State-trace analysis

Single- and dual-process models of memory development

We applied state-trace analysis to determine the number of underlying dimensions or processes needed to explain our developmental data. Following the approach used by Dunn (2008), we plotted the mean proportion of high confidence ("sure-old") responses against the mean summed proportion of high and low confidence ("unsure-old") responses to old-shallow, old-deep and lure items for each age group. This provides a test of the unequal variance signal detection and mixture signal detection models that does not depend upon assumptions concerning the shapes of the memory strength distributions. Whatever the forms of these distributions, these models predict a monotonically increasing state-trace.[5] That is, the plots of confident old responses against low confidence responses for each age and encoding group should be encompassed in a single monotonic function. In contrast, the dual-process model proposed by Yonelinas (1994), assumes that different processes, familiarity and recollection, underlie low and high confidence "old" responses respectively and that recollection is affected more than familiarity by age and depth of encoding (Ghetti & Angelini, 2008). This model therefore generally predicts a non-monotonic state-trace. Different age groups should show plots of high confidence against low confidence old responses that cannot be captured within a single monotonic function. As discussed below, the dual-process model could also predict a monotonic state-trace plot but by only making ancillary assumptions about recollection and familiarity.

The state-trace plot for the current data is given in Figure 3. The Figure is strongly monotonic, with increasing age and encoding depth showing parallel positive effects on high and low confidence old responses. We submitted these data to the test of monotonicity developed by Kalish, Dunn, Burdakov and Sysoev (2016). This test examines whether more than one latent variable is needed to explain the rank order of obtained means across factors like age and encoding depth. The null hypothesis was retained (p = 0.51), indicating that only one latent variable was necessary to account for the recognition data.

In the UVSD and Mixture SD models the nature of this latent variable is clear; it reflects memory strength that increases with both age and encoding depth. Explanation of a monotonic plot from a dual process perspective is considerably more challenging. The dual-process model could explain the monotonic state trace by assuming that the latent variable reflects change in only one of the two hypothesized memory processes (e.g., recollection). This assumption however, is at odds with the results of our dual process modeling, which found age and depth increases in both recollection and familiarity parameters. Likewise, Ghetti and Angelini (2008) found reliable age increases in familiarity between 6- and 10-years of age. Alternately, a dual process approach could explain a monotonic state trace results by assuming that age and depth have parallel effects on recollection and familiarity. But if this is the case then there seems no basis for assuming two separate underlying processes. Hence, the UVSD and Mixture SD models provide a more straightforward and parsimonious explanation of the single latent variable needed to explain age and depth effects on recognition.

Discussion

This study examined whether the development of recognition memory from 6 years of age to adulthood is best explained by a model that posits a single underlying mechanism based on memory strength or a two-process model that posits separate mechanisms of memory strength and conscious recollection. We first carried out a study that was patterned closely on one of the key developmental experiments that has been seen as supporting the dual-process account (Ghetti & Angelini, 2008). We then modeled the developmental data using the dual-process model used in that work, as well as several variants of the single-process approach. Single-process accounts assume that, all things being equal, the memory strength elicited by a previously experienced item, will increase with age. However the underlying process governing recognition remains unchanged. The competing models were compared in terms of their fit to the data and changes in key parameters across age and depth of encoding conditions. They were also compared using a state-trace plot to identify the number of latent variables needed to explain the key trends.

The recognition performance that we observed, as measured by hit and false alarm rates, replicated the major trends reported by Ghetti and Angelini (2008). Hit rates increased from 6-7 to 9-10 years of age and were higher following deep encoding. Larger age changes in hit rates were found following deep than shallow encoding. False alarms were rare and did not vary with age. Retrieval of an episodic feature of the study items (color) also underwent age and encoding changes that were similar to those found by Ghetti and Angelini (2008).

Despite these replications, our modeling of the recognition and confidence data led to a very different set of conclusions from Ghetti and Angelini (2008). Neither the dual-process model nor the simplest single-process account (equal variance signal detection model) provided a satisfactory account of the data in any of the age groups studied. By contrast, in absolute terms, two more complex single-process models (unequal variance signal detection, mixture signal detection model) each provided a good fit to the recognition data for 6-7 year-olds and adults. They also produced an adequate fit to 9-10-year-olds' data, but only when an outlier was removed. Crucially, in relative terms, the UVSD and mixture models provided a better account of the recognition data for all three age groups than the dual-process model.

These conclusions were reinforced by the results of a state-trace analysis of recognition responses. Consistent with the single-process memory strength models, the state-trace showed that only a single latent process was needed to explain the developmental data.

Each of the successful single-process models incorporates an additional parameter that affects recognition responding beyond memory strength (variability in memory strength for targets in UVSD, attention to targets in the Mix SD). Neither of these parameters however, reflects a memory process that is fundamentally distinct from the assessment of memory strength. In fact most instantiations of these models assume that the variability and attention parameters are positively correlated with the measure of memory strength (d'). In this respect the processing assumptions of these models can be seen as quite different to those of the dual-process approach.

In this regard, the parameter estimates of the dual- and single-process models offer very different pictures of how recognition develops. Like Ghetti and Angelini (2008) we found that in the dual-process model, the parameters corresponding to both recollection and familiarity increased with age (especially from 6 years to 9 years) and with depth of encoding. Notably however, in the single-process models which provided a superior fit to the data, age and encoding changes in memory performance were explained primarily by changes in the memory strength parameter.

Based on the current data it is not possible to say with confidence which of the two successful single-process models provides the best account of memory development. The modeling results suggest an advantage for the mixture model over the unequal variance model but the comparative fit indexes suggest that this advantage was relatively small. One way to adjudicate between these models in future work will be to use experimental preparations designed to impact on the theoretical parameters of each model (e.g., variability in memory strength of study items, attention to study items), and to examine whether each model behaves in a manner consistent with these predictions.

The current results raise the general question of the best way to interpret age increases in "memory strength". One interpretation is that the strength parameter reflects the amount of mnemonic information that is retrieved when a probe is presented (Dunn, 2008). Some theorists (e.g., Hautus, Macmillan, & Rotello, 2008; Slotnick & Dodson, 2005) have suggested that such mnemonic information includes both features of the studied item as well as source or contextual information about where and when the item was encountered.

According to this view, on balance, older children will retrieve more item-level and source information about a probe than younger children.

The current modeling and state-trace analyses do not allow us to evaluate these interpretations of memory strength. To do so will require the more systematic manipulation of item and source information and more fine-grained modeling of the process of evidence accumulation in children's memory. What we can say with confidence is that the age changes in recognition performance that we observed can be explained by assuming developmental change on a single underlying psychological dimension rather than separate dimensions of familiarity and recollection.

It should be noted that the superior performance of UVSD and MixSD over the high threshold dual-process model in fitting our developmental data does not necessarily rule out more complex dual-process models of memory. For example, Wixted and Mickes (2010) proposed a continuous dual process model (CDP) in which recognition responses are determined by the sum of signals on two continuous dimensions corresponding to recollection and familiarity. A strength of this approach is that it can explain cases where adults give "know" responses to target items with high levels of confidence or "remember" responses with low confidence. Such cases are problematic for single-process models like UVSD and MixSD as well as for the Yonelinas (1994) high threshold account. Examining the CDP model in the context of memory development however requires a method for eliciting reliable remember and know judgments from young children, a goal that has yet to be fully achieved. When modelling recognition and confidence data without separate remember and know judgments, as was the case in the current study, the CPD reverts to a single process model (i.e., UVSD).

Implications for other work on single- and dual-processes in memory development

Ghetti and Angelini (2008) showed that the age changes in visual recognition memory are consistent with the high threshold dual-process model proposed by Yonelinas (1994). What we have shown however is that such age changes are better explained by versions of a single-process model.

This suggests that other evidence that has been seen as supporting dual-process accounts of memory development should also be re-examined. A behavioral study by Brainerd et al. (2004) used a conjoint recognition procedure in which children aged 5- to 14-years were first presented with lists containing semantically related material (e.g., names of animals such as "cow"). Different groups then made recognition judgments based on different criteria. In the verbatim condition, children were instructed to accept probes only if they were presented at study (as per convention recognition instructions). In the verbatim + gist condition they could accept probes that were presented at study or which had a similar meaning. In the gist alone condition they were asked to accept only items with a similar meaning to the study items but to reject the study items themselves. An algebraic comparison of test responses in each condition led to estimates of the extent to which responding was driven by recollection of study items as opposed to “familiarity” based on semantic content.

These data are interesting and deserve further investigation. However we would make the following points. First, the method assumes that “recollection” and “familiarity-based” processes make independent contributions to individual performance. This assumption has been shown to be problematic (Curran & Hintzman, 1995; Ratcliff, Van Zandt & McKoon, 1995). Second, because the recollection and familiarity-based processes studied by Brainerd et al. are based on perceived semantic similarity to experienced material, it is not clear that they match the processes of “recollection” and familiarity” targeted by Ghetti and Angelini (2008), which involve no explicit reference to the semantic content. Nevertheless, it would be useful in future work to compare the fits of Brainerd et al.’s (2004) dual-process model to conjoint recognition data with fits based on the types of single-process signal detection models studied here.

The current work also has implications for electrophysiological and brain imaging findings that have been seen as evidence for distinct recollection and familiarity processes in memory development (e.g., Mecklinger, Brunnemann, & Kipp, 2010; Ofen, Kao, Sokol-Hessner, Kim, Whitfield-Gabrieli, & Gabrieli, 2007). Although we did not collect such measures we note that analogous work on adult memory has often been shown to be consistent with a single-process interpretation. de Zubicaray, McMahon, Dennis and Dunn (2011), for example, found that a single-process model (UVSD) produced a better fit to the fMRI signals associated with adult recognition responses and memory confidence ratings than the Yonelinas (1994) dual-process model. In a similar vein, Freeman, Dennis and Dunn (2010) applied state-trace analysis to adult electroencephalographic data obtained during a recognition task (also see Staresina, Fell, Dunn, Axmacher and Henson, 2013). The state-trace plot was most consistent with a single-process interpretation. Such studies suggest that it would be useful to apply such analytic methods to imaging and electrophysiological measures collected while children make recognition judgments.

Conclusions

The current study has shown that data that has previously been seen as supporting a dual-process account of memory development can be explained more accurately and parsimoniously by a theoretical account based on a single underlying dimension such as memory strength. The core developmental change suggested by this single-process model is an age-related increase in memory strength, which in our data, was most pronounced from 6 to 9 years of age.

It is certainly possible however, that other factors that are not fully captured by the single-process models examined in this study contribute to developmental change in memory. For example, there is evidence that the ability to integrate the item-level and contextual features that make up a to-be-remembered episode increases substantially from early to mid-childhood (e.g., DeMaster & Ghetti, 2013; Yim, Dennis & Sloutsky, 2013). Moreover as noted earlier, children's ability to monitor and report their subjective memory experiences also develops markedly over this period. Our results suggest that future work directed towards the study of such factors will likely lead a better understanding of developmental continuities and discontinuities in memory than will the search for distinct age changes in recollection and familiarity.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Bamber, D. (1979). State-trace analysis: A method of testing simple theories of causation. *Journal of Mathematical Psychology, 19*, 137–181.

Berch, D. B., & Evans, R.C. (1973). Decision processes in children's recognition memory. *Journal of Experimental Child Psychology, 16*, 148-164.

Billingsley, R. L., Smith, M-L., & McAndrews, M. P. (2002). Developmental patterns in priming and familiarity in explicit recollection. *Journal of Experimental Child Psychology, 82*, 251-277.

Brainerd, C. J., Holliday, R. E., & Reyna, V. F. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child Development, 75*(2), 505-522.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer-Verlag

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 531-547.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65,* 171-237.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109(4),* 710–721.

DeMaster, D. M., & Ghetti, S. (2013). Developmental differences in hippocampal and cortical contributions to episodic retrieval. *Cortex, 49*, 1482-1493.

de Zubicaray, G. I., McMahon, K. L., Dennis, S., & Dunn, J. C. (2011). Memory strength effects in fMRI studies: A matter of confidence. *Journal of Cognitive Neuroscience, 23,* 2324-2335.

Dunn, J. C. (2008). The dimensionality of the Remember-Know task: A state-trace analysis. *Psychological Review*, *115*, 426-446.

Freeman, E., Dennis, S., & Dunn, J. C. (2010). An examination of the ERP correlates of recognition memory using state-trace analysis. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 97-102.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. E. Tulving, E. Fergus, & I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 229-244). New York, NY: Oxford University Press.

Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child Development*, *79*(2), 339-358.

Ghetti, S., & Lee, J. (2011). Children's episodic memory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 365-373.

Ghetti, S., Mirandola, C., Angelini, L., Cornoldi, C. & Ciaramelli. E. (2011). Development of subjective recollection: Understanding of and introspection on memory states. *Child Development*, *82*, 1954-1969.

Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology, 38*(5), 705-718.

Hautus, M., Macmillan, N. A., & Rotello, C. M. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review, 15*, 889–905.

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

Kalish, M. L., Dunn, J. C., Burdakov, O. P., & Sysoev, O., (2016). A statistical test of the equality of latent orders. *Journal of Mathematical Psychology, 70*, 1-11.

Koenig, L., Wimmer, M. C., & Hollins, T. J. (2015). Process dissociation of familiarity and recollection in children: Response deadline affects recollection but not familiarity. *Journal of Experimental Child Psychology, 131*, 120-134.

Mecklinger, A., Brunnemann, N., & Kipp. K. (2010). Two processes for recognition memory in children of early school age: An event-related potential study. *Journal of Cognitive Neuroscience*, *23*, 435-446.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Newell, B. R., Dunn, J. C., & Kalish, M. L. (2011). Systems of category learning: Fact or fantasy? *The Psychology of Learning and Motivation*, *54*, 167-215.

Ofen, N., Kao, Y-C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S. & Gabrieli, J. D. E. (2007). Development of the declarative memory system in the human brain. *Nature Neuroscience*, 10, 1198-1205.

Pazzaglia, A. M., Dube, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*, 1173-1203.

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology*, *59*, 516 – 548.

Piolino, P., Hisland, M., Ruffeveille, I., Matuszewski, V., Jambaqué, I., & Eustache, F. (2010). Do school-age children remember or know the personal past? *Consciousness and Cognition*, *16*, 84-101.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472-491.

Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, *17*, 78.

Pratte, M. S., & Rouder, J. N. (2012). Assessing the dissociability of recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1591-1607.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http: //www.R-project.org

Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 365-377.

Ratcliff, R., Sheu, C., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*(3), 518-535.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single-process theories and recognition memory. *Journal of Experimental Psychology: General, 124*, 352-374.

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

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single process) model of recognition memory and source memory. *Memory & Cognition*, *33*, 151-170.

Staresina, B. P., Fell, J., Dunn, J. C., Axmacher, N., & Henson, R. N. (2013). Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 3119-3124

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*5, 1379.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*(1), 152-176.

Wixted, J. T. & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*, 1025-1054.

Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory. Items, contexts and relations. *Psychological Science, 24*, 2163-2172.

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.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*(6), 1341-1354.

Yonelinas, A. P., & Jacoby, L. L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions and new directions. *Memory and Cognition, 40*, 663-680.

Footnotes

1 We contacted the senior author of Ghetti and Angelini (2008) about the possibility of fitting these models to the original data. Unfortunately however the data were not available.

2 An additional manipulation was built into the study/test lists. Half of the study pictures were "small" (3 x 3cm) and the other half were "large" (6 x 6 cm). At test half the targets were shown in the same size (congruent) and half were shown in the opposite size (incongruent). Lures were shown in either small or large size, determined randomly. This manipulation was based on Rajaram (1996) who found that targets presented in congruent size had a higher hit rate, which was attributed to higher levels of recollection. We also found a higher hit rate for size congruent targets (M = 0.66, SD = 0.12) than incongruent items (M = 0.64, SD = 0.12), $F(1, 93) = 7.15$, p =.009, η_p^2 = .07. This effect did not interact with level of encoding or age (F's<1.7). The four target models were fit to recognition responses to congruent and incongruent items (aggregated over level of encoding). The model fit results closely paralleled those of the deep shallow modelling, with the UVSD and MixSD models providing a significantly better overall fit after adjustments for complexity than any other model. The full details are available from the authors.

3 Ghetti and Angelini (2008) used a six-point confidence scale (old/new – "very sure", "somewhat sure", "not at all sure"). However pilot testing with a separate group of 6-7 year-olds showed that many did not reliably use the middle point of this scale. We therefore opted for a binary confidence scale.

4. Given that the chi-square test on the G^2 statistic is known to be highly sensitive to small deviations between predicted and observed data we adopted a conservative α rate of .01.

5. With the proviso that in the UVSD model s, the standard deviation of the old item distribution, is approximately constant. The parameter estimates in Table 4 suggest that this was the case.

Appendix A

Formal versions of Recognition Memory Models

Dual-Process models

High Threshold Signal Detection (HTSD, Yonelinas, 1994)

$$f(c) = \Phi(-c) \qquad (1)$$

$$h\,(c) = r + (1\text{-}r)\,\Phi\,(d' - c) \qquad (2)$$

c = is a decision criterion; $f\,(c)$ = probability of making a false alarm; $h\,(c)$ = probability of making a hit; r = the probability that some information from the study episode is recollected; Φ is the normal cumulative distribution function which returns the area under the normal curve to the left of its argument; d' = the difference in familiarity between target and lure distributions

Single-Process models

Equal Variance Signal Detection Model (EVSD)

$$f\,(c) = \Phi\,(-c) \qquad (1)$$

$$h\,(c) = \Phi\,(d' - c) \qquad (2)$$

d' = the difference between the means of the target and lure distributions

Unequal Variance Signal Detection Model (UVSD, Wixted, 2007)

$$f\,(c) = \Phi\,(-c) \qquad (1)$$

$$h\,(c) = \Phi\,((d' - c)/\mathrm{s}) \qquad (2)$$

d' = the difference between the means of the target and lure distributions in units based on the lures; s = the standard deviation of the target distribution

Mixture Signal Detection Model (MixSD, DeCarlo, 2002)

$$f\,(c) = \Phi\,(-c) \qquad (1)$$

$$h\,(c) = (1 - \lambda)\,\Phi\,(-c) + \lambda\,\Phi\,(d' - c) \qquad (2)$$

d' = the difference between the means of the target and lure distributions; λ = the probability of paying attention to a target at study

Appendix B

Revised model fits of 9-10-year-old data with outlier removed (n= 31)

Memory model parameters and goodness of fit measures

Age	Model[a]	Model parameters for encoding depth conditions [b]								Decision Criteria				Goodness of Fit		
		Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep							
		d'	*d'*	*r*	*r*	*s*	*s*	*λ*	*λ*	*c1*	*c2*	*c3*	*LL*[c]	G^2	*df*	*p*
9-10-year-olds	MixSD	1.85	2.47					0.69	0.91	1.63	1.27	0.72	3647.47	4.84	2	0.089
	UVSD	1.35	2.50			1.39	1.39			1.63	1.25	0.722	3649.55	9.01	2	0.011
	HTSD	0.93	1.65	0.23	0.45					1.62	1.23	0.73	3652.29	14.49	2	0.001*
	EVSD	1.25	2.10							1.50	1.20	0.75	3662.09	34.09	4	0.000*

[a] MixSD = Mixture signal detection model, UVSD = Unequal variance signal detection model, HTSD = Dual process, high-threshold signal detection model, EVSD = Equal variance signal detection model

[b] *d'* = Memory strength; *r* = Recollection, *s* = Variability of the target distribution, *λ* = Probability of attending to study item

[c] LL = Negative log likelihood of fitted model

* denotes a significant discrepancy between the ideal model and data at α =.01

Table 1.

Recognition responses in each age and encoding condition

	Hit Rate		False Alarm Rate
	Shallow	Deep	
6-7 years	0.46 (0.12)	0.69 (0.14)	0.11 (0.10)
9-10 years	0.53 (0.15)	0.83 (0.11)	0.10 (0.08)
Adults	0.56 (0.17)	0.81 (0.12)	0.12 (0.09)

Note. Standard deviations are shown in parentheses.

Table 2.

Mean proportion of "sure" responses for hits and false alarms

	Hits		False Alarms
	Shallow	Deep	
6-7 years	0.09 (0.07)	0.49 (0.10)	0.06 (0.07)
9-10 years	0.11 (0.08)	0.56 (0.17)	0.05 (0.05)
Adults	0.15 (0.06)	0.54 (0.14)	0.05 (0.04)

Note. For hits at each level of encoding the proportion given was computed as the number of "sure-old" responses over the total hits. For false alarms the proportion was the number of "sure-old" responses over the total false alarms. Standard deviations are shown in parentheses.

Table 3.

Mean accuracy of recall of color of old items in each age and encoding condition

	Shallow	Deep
6-7 years	0.66* (0.12)	0.56* (0.13)
9-10 years	0.60* (0.16)	0.66* (0.19)
Adults	0.58* (0.15)	0.83* (0.15)

Note. Standard deviations are shown in parentheses.

* Denotes significant at p<.01 or better

Table 4

Memory model parameters and goodness of fit measures

Age	Model[a]	Model parameters for encoding depth conditions [b]								Decision Criteria				Goodness of Fit		
		Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep							
		d'	*d'*	*r*	*r*	*s*	*s*	*λ*	*λ*	*c1*	*c2*	*c3*	*-LL*[c]	G^2	*df*	*p*
6-7-year-olds	MixSD	1.64	2.17					0.64	0.79	1.54	1.22	0.72	3985.18	3.96	2	0.138
	UVSD	1.10	1.90			1.32	1.50			1.54	1.21	0.72	3986.46	6.52	2	0.038
	HTSD	0.79	1.13	0.17	0.38					1.53	1.20	0.73	3988.75	11.11	2	0.004*
	EVSD	1.06	1.62							1.42	1.16	0.75	4000.66	34.93	4	0.000*
9-10-year-olds	MixSD	1.83	2.44					0.70	0.91	1.66	1.27	0.70	3749.46	11.09	2	0.004*
	UVSD	1.34	2.42			1.36	1.34			1.65	1.26	0.70	3755.42	17.06	2	0.000*
	HTSD	0.95	1.73	0.22	0.39					1.63	1.23	0.71	3752.44	23.03	2	0.000*
	EVSD	1.23	2.00							1.49	1.18	0.71	3765.17	42.51	4	0.000*
Adults	MixSD	1.91	2.44					0.67	0.90	1.71	1.17	0.51	3857.55	2.45	2	0.294
	UVSD	1.36	2.46			1.47	1.47			1.72	1.15	0.51	3859.97	7.29	2	0.026
	HTSD	0.86	1.50	0.25	0.47					1.70	1.12	0.52	3865.76	18.86	2	0.000*
	EVSD	1.22	2.00							1.52	1.09	0.55	3883.488	54.32	4	0.000*

[a] MixSD = Mixture signal detection model, UVSD = Unequal variance signal detection model, HTSD = Dual process, high-threshold signal detection model, EVSD = Equal variance signal detection model

[b] *d'* = Memory strength; *r* = Recollection, *s* = Variability of the target distribution, *λ* = Probability of attending to study item

[c] -LL = Negative log likelihood of fitted model

* denotes a significant discrepancy between the ideal model and data at α =.01

Table 5

AIC and BIC measures of relative model fit

Age	Model	Fit Measures	
		AIC	BIC
6-7-year-olds	MixSD	7984.35	8028.58
	UVSD	7986.91	8031.14
	HTSD	7991.50	8035.72
	EVSD	8011.33	8042.92
9-10-year-olds	MixSD	7512.91	7557.14
	UVSD	7518.88	7563.10
	HTSD	7524.84	7569.07
	EVSD	7540.33	7571.92
Adults	MixSD	7729.10	7773.10
	UVSD	7733.946	7777.948
	HTSD	7745.52	7789.52
	EVSD	7776.975	7808.405

Figure 1.

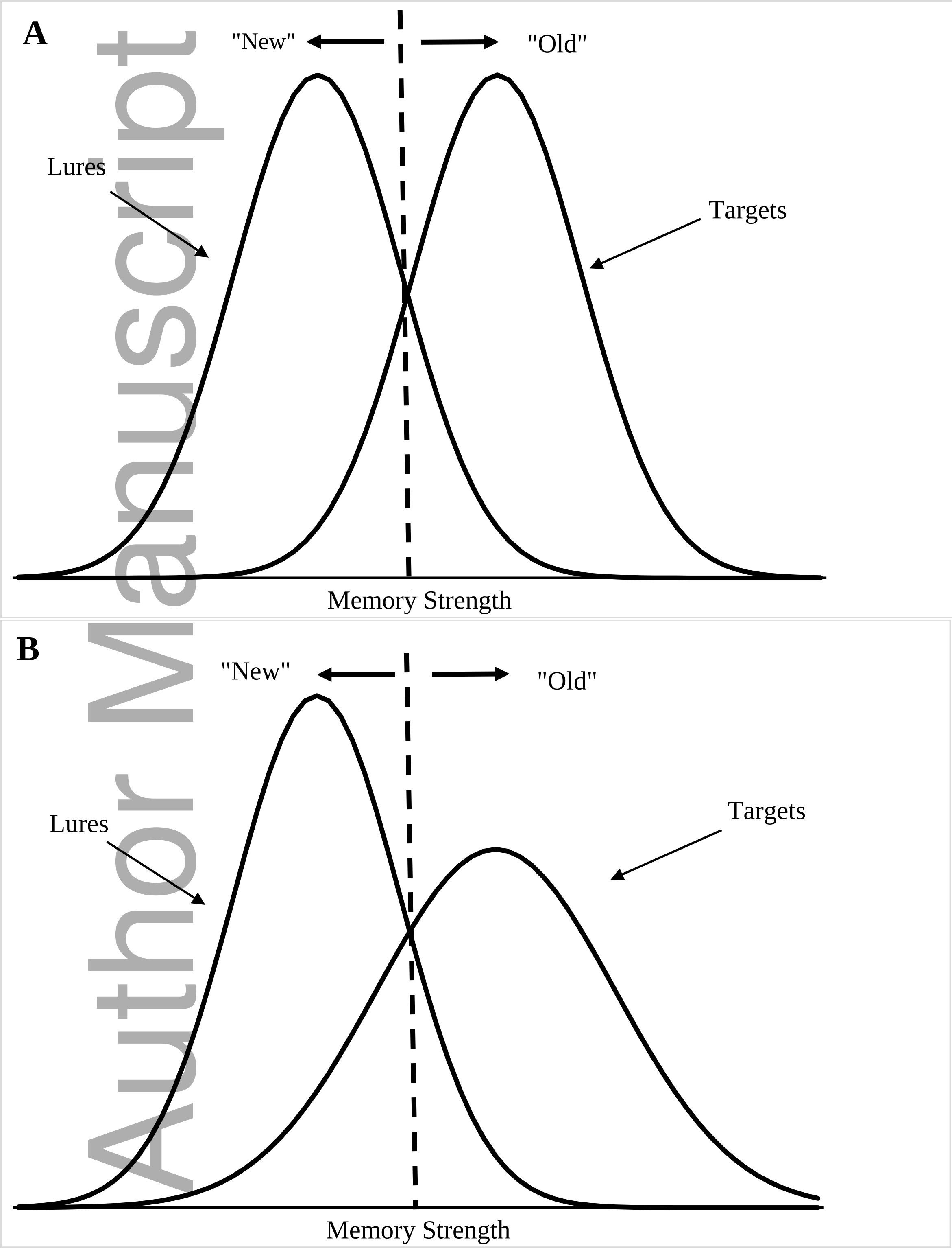
A
"New"
"Old"
Lures
Targets
Memory Strength
B
"New"
"Old"
Lures
Targets
Memory Strength

Figure 2

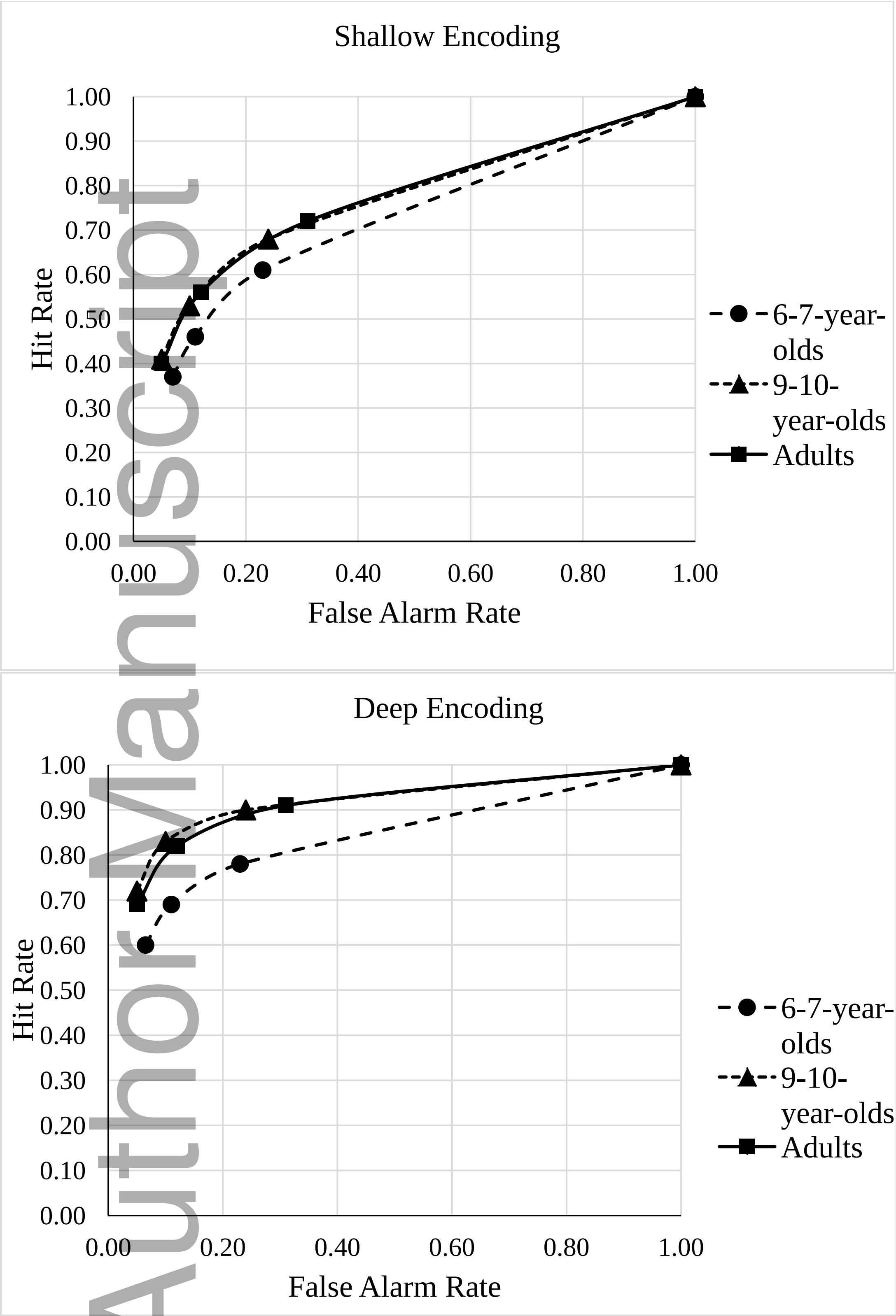
Shallow Encoding
Hit Rate
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.20
0.40
0.60
0.80
1.00
False Alarm Rate
6-7-year-olds
9-10-year-olds
Adults
Deep Encoding
Hit Rate
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.20
0.40
0.60
0.80
1.00
False Alarm Rate
6-7-year-olds
9-10-year-olds
Adults

Figure 3

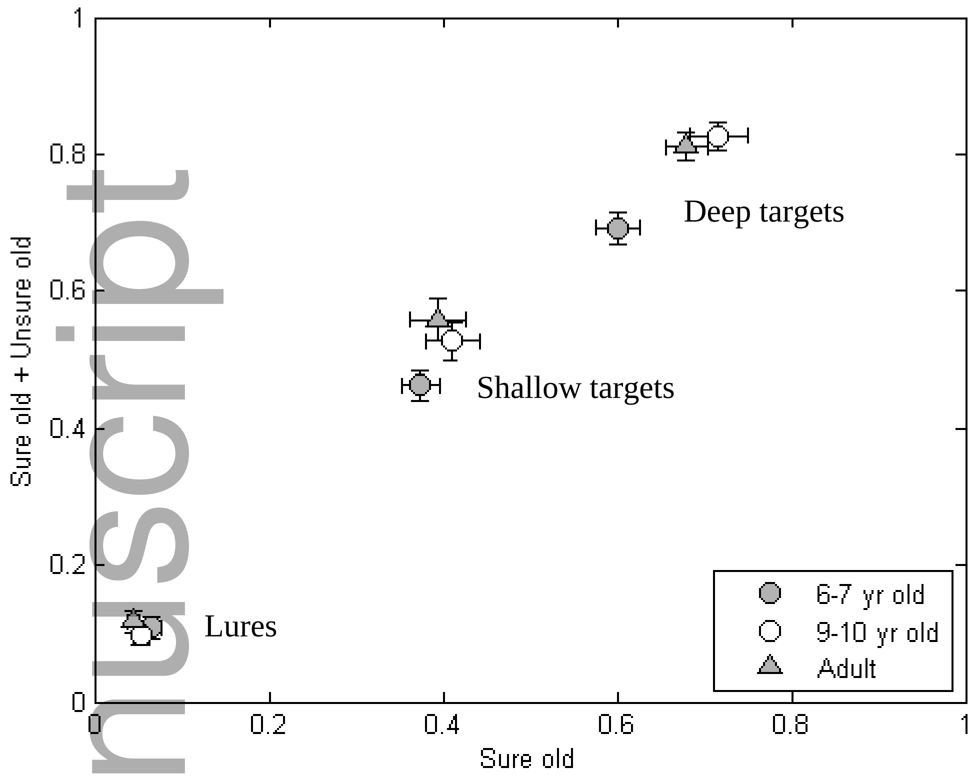

Figure Captions

Figure 1. Equal variance (panel A) and unequal-variance (panel B) signal detection models of recognition memory. The dashed line represents the subjective criterion for saying "old"

Figure 2. ROC curves for each encoding and age condition based on observed recognition responses.

Figure 3. State-trace plot for "old" responses to lures, old shallow targets and old deep targets. Error bars indicate standard errors.