Context and Category Information in Children and Adults

Adam F. Osth (adamosth@gmail.com)
Center for Cognitive Science
The Ohio State University
208F Ohio Stadium East, 1961 Tuttle Park Place
Columbus, OH 43210, USA

Simon Dennis (simon.dennis@gmail.com)
Memory and Language Lab
The Ohio State University
225 Psychology Building, 1835 Neil Avenue
Columbus, OH 43210, USA

Vladimir M. Sloutsky (sloutsky.1@osu.edu)
Center for Cognitive Science
The Ohio State University
208C Ohio Stadium East, 1961 Tuttle Park Place
Columbus, OH 43210, USA

Abstract

An experiment by Dennis and Chapman (in press) found that as the length of a categorized list of materials increased, the false alarms to unrelated distracters decreased, a finding suggesting that adults are best described by context-noise models of recognition memory. Developmental evidence demonstrating that children the age of five are more sensitive to item information suggests that children might be described by item-noise models. We tested children and adults' performance and eye movements during recognition and found that adults' usage of category context was evident in both their performance and in their eye movements. Children, however, did not give conclusive evidence in their memory performance but their eye movements did not reflect usage of category context.

Keywords: Recognition memory; Inverse list length effect; Categorization and memory; Development of memory; REM; BCDMEM; context-noise; item-noise

Introduction

Current models of recognition memory, such as the REM model (Shiffrin & Steyvers, 1997) and the BCDMEM model (Dennis & Humphreys, 2001) are capable of making accurate predictions about a number of previously problematic effects in the literature, such as the list-strength effect (Ratcliff, Clark, and Shiffrin, 1990) and the mirror effect (Glanzer and Adams, 1985). However, these models not only possess different architectures but also capture the same trends in the data using different sources of information and interference.

The REM model assumes that during the study phase, each item is stored as a separate, noisy representation. During the test phase, a probe item is compared against every item in memory and an activation value is calculated based on the degree of match between each studied item and the presented probe item. These activation values are then averaged and a mean activation value that is sufficient (exceeding a fixed criterion value) produces a yes response. If distracters happen to have sufficient match with some of the studied items in memory, this produces a false alarm to the distracter. The REM model, as well as other global-matching type models (see Clark & Gronlund (1996) for a review), are classified as item-noise models due to the fact that interference is produced by the content and number of studied items.

BCDMEM, in contrast, assumes that during the study phase, each item isn't stored but is instead bound to the study context. During the test phase, the probe item cues all previous contexts in which the item was studied in, including learned elements of the study context. Additionally, the context of the study episode is reinstated. This reinstated context is then compared against the retrieved contexts of the item to evaluate whether or not the item was presented during the experiment, and a sufficient degree of match between the elements of the study context and matching elements in the retrieved contexts produces a yes response. If a distracter item happens to have been experienced in a large number of contexts, such as a high frequency word in the English language, then it is more likely for the retrieved context layer to spuriously contain elements of the study context and a false alarm can be produced. Consequently, this model is classified as a context-noise model because previous contexts are the principal source of interference.

Because these models account for the same effects using different sources of information and interference, determining which model is correctly representing the memory system requires looking at more current evidence in the literature. Dennis and Chapman (in press) recently found that when list length was varied between 10 and 80 items but the number of categories was kept constant (essentially varying the number of exemplars per category), false alarms to unrelated distracter items were lower in the long list relative to the short list. This qualitative trend was dubbed the 'inverse list length effect,' which is the opposite of the predictions of the REM model. REM predicts a slight increase to unrelated distracters with increased list length, as extra study items produce additional noise during recognition. However, this effect can be handled by the BCDMEM model with the assumption that when participants are reinstating the context of the study episode, the categories learned while studying the long list become a part of the reinstated study context. These added elements of category context are then matched to the category contexts cued by the items, producing higher likelihoods of a "yes" response for matching category items and lower likelihoods of "yes" responses for mismatching category items. The authors further argue that it would be impossible for the REM model or virtually any other model that solely relies on information from individual exemplars to capture this effect.

There exist developmental evidence, however, that suggest that the source of interference may change through development. Sloutsky and Fisher (2004) conducted an experiment in which participants, which consisted of adults and children the age of five, either participated in a categorization task in which they had to induce a novel category property to animal photos, or they participated in a baseline condition in which they merely studied the photos. A surprise recognition task followed either of the two conditions, and it was found that adults experienced a sharp decrement in their ability to discriminate between studied and non-studied category items due to an increase in false alarms to related distracters. Children, in contrast, experienced no such decrement in their memory performance between the two conditions. The authors attribute this dissociation in performance between the two age groups to the fact that adults are much more sensitive to category-based information, while children are much more sensitive to item-based information.

These findings are consistent with a large number of findings from the research review performed by Brainerd, Reyna, and Ceci (2008). In this review, a number of experiments were discussed in which it was demonstrated that children are much less susceptible to false memory errors, a pattern which increases with age up until adulthood. These effects and age trends were evident in paradigms ranging from DRM type tasks to suggestibility experiments. The authors attribute these errors, particularly the decreased likelihood of recalling or falsely recognizing items highly similar to items presented at test, to children’s weaker ability to spontaneously extract gist from the test materials in the same manner as adults.

We argue that if children are indeed more focused on item information and are weaker in their ability to extract gist during the study phase, then they should be deficient in their ability to extract and reinstate category context in the same manner as adults. We tested this by running both adults and children in a paradigm very similar to that used by Dennis and Chapman (in press). If our hypothesis about children's inability to use context is correct, then they should be unable to exhibit an inverse list length effect and their performance may follow the predictions of item-noise models of recognition. Adults, in contrast, should behave consistently with previous findings and experience facilitation in their ability to reject unrelated distracters in the long list condition.

Experiment 1

Method

Participants Participants were 65 children (33 female and 32 male, M = 4.87 years, SD = 0.61 years) and 83 adults (36 women and 47 men, M = 19.7 years, SD = 2.93 years). Child participants were recruited from suburbs in Columbus, OH. Adult participants consisted of undergraduate students from The Ohio State University participating for course credit.

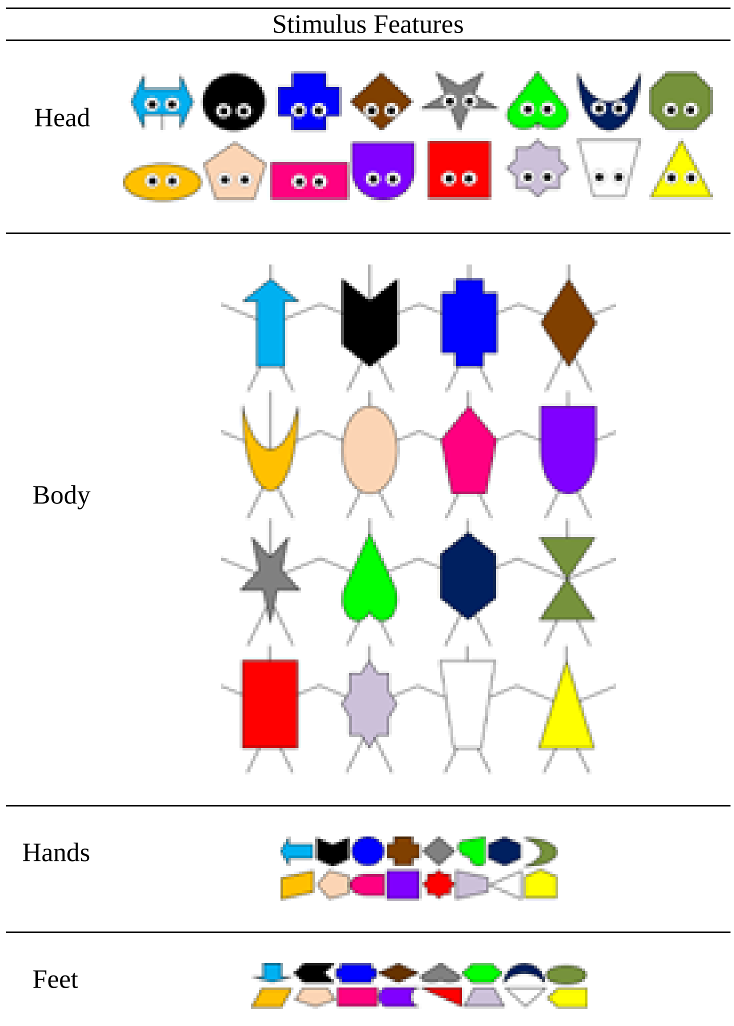

Figure 1: All stimulus feature possibilities in exemplar construction. For the sake of space efficiency, only left hands and left feet are presented.

Stimuli Visual stimuli consisted of artificial creatures that were composed of four different body parts: a head, a body, a pair of hands, and a pair of feet. Each body part was composed of a unique color and shape pairing. The creatures were assembled by randomly selecting a shape and a color for each component from a selection of over 16 different colors (common to all body parts) and 16 different shapes (unique to each body part). All shape and color features are detailed in Figure 1.

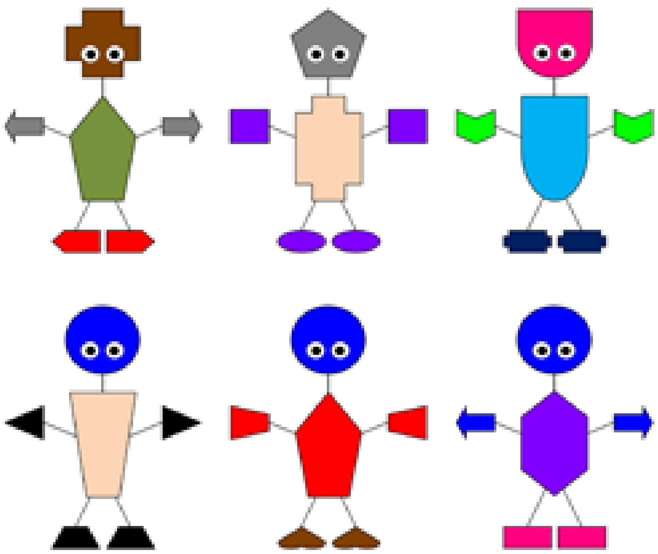

Figure 2: Examples of composite figures created from the sample features. *Top row:* Examples of three non-categorized exemplars. *Bottom row:* Examples of three categorized exemplars potentially seen in the long list.

For the study phase of the short list, a total of 8 exemplars were presented to the participant with unique shapes and colors for the different components (head, body, hands, and feet). For the study phase of the long list, 64 exemplars were presented and blocked into 8 categories of 8 exemplars each. Categories were defined as exemplars which shared heads of the same shape and color while the hands, feet, and body of the shared category exemplars consisted of different shape and color combinations. All categories exhibited unique, non-overlapping shape and color combinations for the heads. For the body, hands, and feet, all categories used the same colors and shapes (all randomly selected from the 16 available), no shape/color combinations were reused between categories and no shapes or colors were used for more than one exemplar within a given category. It should also be mentioned that the same number of features were used in both the short and long list conditions. Examples of possible exemplars created for short lists and long lists can be seen in Figure 2.

Distracter items in the test list were divided into two types: related distracters and unrelated distracters. Related distracters were constructed by selecting one head from each of the 8 presented categories and using reshuffled combinations of previously presented shapes and colors for the body, hands, and feet. Unrelated distracters were constructed in the same manner as related distracters, except that the heads were not categorically related to the presented stimuli and instead were composed of new combinations of shapes and colors that were not presented at test (all remaining colors and shapes not sampled for the study items). It should be noted that all shapes and colors have an equal probability of being used in target, related distracter, and unrelated distracter items. Examples of possible distracter item composites can be seen in Figure 3.

The stimuli were presented on the center of the computer screen. All stimuli and tasks in the experiment were controlled by E-Prime 2.0 Professional software.

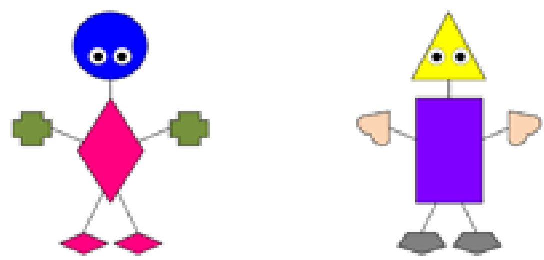

Figure 3: Examples of possible distracter items. *Left:* a related distracter item, featuring an identical head as the above category exemplars. *Right:* an unrelated distracter, featuring a head composed of non-presented features.

Procedure Upon arrival, participants were randomly selected into either the short list condition or the long list condition. Participants were then briefed about the stimuli they would be viewing. They were given incidental learning conditions, in that they were told to look at the stimuli and make a decision as to whether each stimulus was scary or funny, but were not told that they would be tested on their memory later in the experiment. During the study phase of the experiment, each exemplar was presented on the screen for 4500 ms while evaluating the exemplar to make the decision described above. Adult participants gave their responses by pressing keys on the keyboard while child participants gave their responses verbally to the experimenter, who then recorded the response on the keyboard. Each exemplar was preceded by a fixation cross which appeared for 500 ms.

To control for the different retention intervals between the short and long list, after completing the study phase of the experiment, both the short and long lists were followed by a distracter task that took place for 340 seconds for the short list and 60 seconds for the long list. The distracter task consisted of a rhythm game where participants listened to a sequence of 4 drumbeats and were then asked to tap out the sequence on the spacebar of the keyboard at the same tempo that the beats had played.

After completing the distracter task, participants were then instructed that they would be tested on their memory of the items from the study condition. Participants were instructed to respond "yes" if they recognized the item from the study phase, or to respond "no" if they did not recognize the item. The adult participants were instructed to give their responses on the keyboard while the children gave verbal responses to the experimenter who recorded the responses on the keyboard.

The test list consisted of 24 items: 8 of which were target items drawn from the study list, 8 of which were related distracters, and 8 of which were unrelated distracters. For the long list, each target item and related distracter was selected from each of the separate categories with no category being sampled more than once, such that all categories were represented on the test list.

Adult participants were tested in a laboratory at the university. Child participants were tested in local daycares or preschools by trained adult experimenters.

Results and Discussion

d' scores were calculated as a measure of memory sensitivity for all participants. Since we were most principally interested in the usage of the head information in making recognition judgments, only hits and false alarms to unrelated distracters were used in the calculations. Edge corrections were performed by adding 0.5 to the hit and unrelated false alarm counts and 1 to the target and unrelated distracter counts, as hit rates of 1.0 or false alarm rates of 0.0 produce infinite values for *d'* (Snodgrass & Corwin, 1988). Participants with *d'* less than or equal to 0 were excluded from the analysis (7 children and 4 adults). Hits, false alarms to related distracters, and false alarms to unrelated distracters are summarized in Table 1.

Table 1: Mean Proportions of Hits, Related False Alarms (FA), Unrelated False Alarms, and Mean *d'* Scores

	Adults		Children	
	Short List	Long List	Short List	Long List
Hits	0.82	0.86	0.66	0.64
Related FA	0.47	0.79	0.52	0.60
Unrelated FA	0.20	0.16	0.22	0.22

A repeated measures analysis of variance (ANOVA) was conducted on the responses for each subject, with list condition (short vs. long) and age group (adults vs. children) as between subjects factors and item type (target vs. related distracter vs. unrelated distracter) as a within subjects factor. Results indicated a significant main effect of item type, $F(2, 405) = 132.50$, $p < .001$, list type, $F(2, 405) = 4.28$, $p < .05$, as well as a significant item by age interaction, $F(2, 405) = 10.48$, $p < .001$, an age by list interaction, $F(1, 405) = 5.34$, $p < .05$, and an item by age by list interaction, $F(2, 405) = 3.09$, $p < .05$.

Planned post-hoc comparisons revealed a significant difference in adults' related false alarm rates across the two conditions (t = 24.05, $p < .001$), a difference which was insignificant for children (t = 2.01, $p > .05$). This replicates the findings of Sloutsky and Fisher (2004) but does not distinguish between the two models of recognition memory we're comparing against. Thus, planned post-hoc comparisons were also calculated on the differences in unrelated false alarms between the two list conditions for both adults and children, and revealed insignificant differences for both age groups (adults: $t = 1.16$, children: $t = .04$, ps > .05). Thus, neither age group revealed an inverse list length effect in their memory performance. However, there were differences in the groups' reaction times.

Table 2: Mean of Median RTs for Target Items, Related Distracters, and Unrelated Distracters

	Adults		Children	
	Short List	Long List	Short List	Long List
Target	1009	1011	3222	3126
Related	1344	1053	3130	3137
Unrelated	1198	944	2944	2859

To counteract positive skew in reaction time data, we took the median of each participant's reaction times for each of the 8 trials of every item type. The means of these median RTs can be seen in Table 2. We subjected the RTs to a repeated measures analysis of variance with list condition and age group as between subjects factors and item type as a within subjects factor. Results indicated a significant main effect of age, $F(2, 405) = 235.4$, $p < .001$. Planned post-hoc comparisons were also calculated on the RT differences between the two conditions for each item type and for each age. Significant differences between the two list conditions were found in adults for related distracters ($t = 2.95$) and unrelated distracters ($t = 3.75$), ps < .01. No significant differences in reaction times were found between the two list conditions in children, however it should be mentioned that children's responses were recorded by an experimenter and are thus difficult to interpret.

Because adults in the long list are quicker to react to unrelated distracters without receiving any decrement in accuracy, it is clear that they are in fact receiving a facilitation in rejecting unrelated distracters, which is not only contrary to the predictions of item-noise models but in accordance with the context-noise approach. However, this evidence does not give us any indication as to which model describes children's memory judgments. For this, we decided to run the same experiment in an eye tracker as a way of measuring the information that is being used in the test phase. Considering that the head is the most relevant feature of an unrelated distracter, the clearest prediction that can be made is that participants that have successfully abstracted the category context in the long list will look significantly longer at the head relative to participants in the short list condition.

Experiment 2

Method

Participants Participants were 34 children (15 female and 19 male, *M* = 5.25 years, SD = 0.60 years) and 43 adults (18 female and 25 male, *M* = 19.5 years, SD = 1.29 years), participated in this experiment. Child participants were recruited from suburbs in Columbus, OH. Adult participants consisted of undergraduate students from The Ohio State University participating for course credit.

Apparatus Eye gazes were measured using a Tobii T60 eye tracker with a sampling rate of 60 Hz (60 data points collected per second). The device is integrated into a 17-inch monitor within a testing booth. A camera adjacent to the eye tracker provided a live feed to a trained experimenter at a nearby computer, who was able to monitor both the participant's eye movements as well as the stimuli they were viewing.

Procedure The procedure was nearly identical to that of Experiment 1 with the exception of a couple of minor adaptations to make this experiment compatible with the usage of an eye tracker. Each stimulus component (head, body, hands, and feet) was given a pre-determined area of interest (AOI) for recording eye gaze movements. To keep participants visual attention, we used a gaze-contingent fixation point between all trials in both the study phase and the test phase such that a stimulus would only be presented if participants maintained their gaze on the fixation point for a randomly calculated time interval between 300 and 700 ms. Additionally, since the stimulus appears in the center of the screen, the fixation point was randomly presented in the center of one of four quadrants on the screen to ensure that first looks were not biased by the fixation position.

All participants, including both adults and children, were run by trained adult experimenters at a nearby computer throughout the duration of the experiment. Because trials continue until a response is given, allowing adults to enter their own responses on a keyboard and having children's responses be entered by an experimenter will yield different patterns of data for the two groups. To make the data comparable between the two age groups, both adults and children gave all responses to the experimenter who entered them on a keyboard.

Results and Discussion

d' scores were calculated in the same manner as Experiment 1 and all participants with d' scores less than or equal to 0 were excluded from the analysis (2 adults and 4 children).

Because the head was the category relevant feature, all analyses were restricted to that area of interest. The dependent measure we selected was the proportion of looks at the head, which was calculated for each item type. These results can be seen in Table 3. Because trials continued until participants gave their responses, with some trials continuing for as long as 10 seconds, the calculation was restricted from the start of the trial up until the mean reaction time (2 seconds for adults and 3 seconds for children).

Table 3: Mean Proportion of Looks at the Head for Target Items, Related Distracters, and Unrelated Distracters

	Adults		Children	
	Short List	Long List	Short List	Long List
Target	.261	.282.	.272	.247
Related	.272	.256	.276	.259
Unrelated	.261	.345	.296	.369

Considering that there is no category information in the short list (there were no repetitions of category exemplars), we interpreted looks at the head in the short list as a baseline degree of looking when only item information is available. Since the test phase is identical in both list conditions, any increase in looking at the head in the long list above that of the short list has to be due to differences in the study phase, most notably the repetitions of category exemplars. Visualizations of the looks at the head over time can be found in Figures 4 and 5.

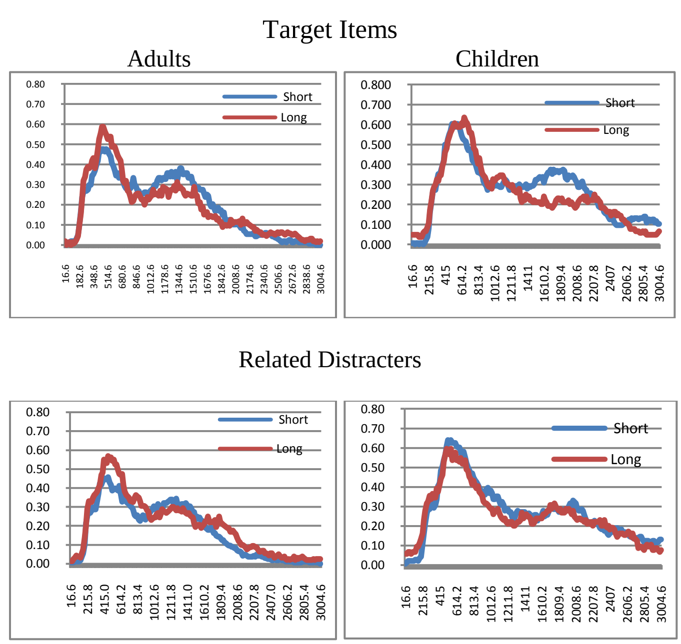

Figure 4: Differences in proportions of looking at the head target and related distracter presentations for both children and adults recorded at each refresh rate (every 16.6 ms).

T-tests were calculated on the differences between the two list conditions for each item type to determine if the differences in looking were significant. For adults, differences between the two lists were insignificant for both target items ($t(39) = .90$, $p > .05$) and related distracters ($t(39) = -.53$, $p > .05$). This is not surprising, considering that for both target and related distracter items, the head is not diagnostic of whether or not the item was on the list. This is not the case for unrelated distracters, where the category information (i.e.: the head) is the most relevant feature for discrimination. A t-test between the two list conditions for unrelated distracters was significant ($t(39) = 2.39$, $p < .05$), in that adults looked significantly longer at the head in the long list condition relative to the short list condition.

For children, differences between the two list conditions for target items ($t(28) = -.76$, $p > .05$), related distracters ($t(28) = -.46$, $p > .05$), and most importantly, unrelated distracters ($t(28) = 1.97$, $p > .05$) were all insignificant. Because children in the long list condition were not using the category relevant information above and beyond that of the short list, we interpret this to mean that category context information was not being accessed above and beyond that of item information. However, considering that this p-value is close to the significance margin, it is possible that there are subsets of children showing the effect that are outnumbered by children not showing the effect.

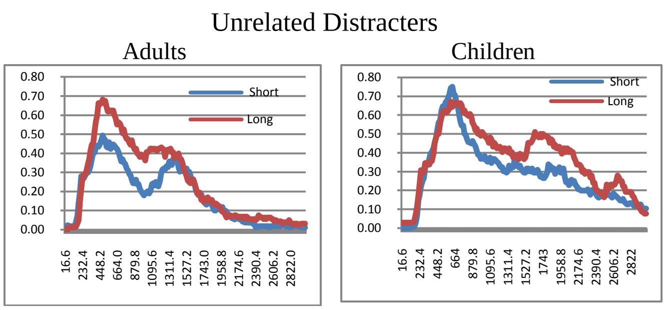

Figure 4: Differences in proportions of looking at the head unrelated distracter presentations for both children and adults recorded at each refresh rate (every 16.6 ms).

Conclusions

To summarize, for adults, increasing category length not only facilitated the rejection of non-category items but this facilitation manifested itself in a bias for category relevant features in their eye movements. For children, no such facilitation could be found in either their behavioral data or in their eye movements, implying that they may be meeting the predictions of the item-noise models of recognition memory.

We believe this is important research because despite there being a large volume of research on developmental differences in episodic memory, there has been little work connecting these differences to the components and processes of current memory models. We hope that this work as well as future work will make clear connections between the developmental literature and the modeling literature and use them to construct a detailed theory of how memory changes with development.

Acknowledgments

This research has been supported by grants from the grants from the NSF (BCS-0720135), from the Institute of Education Sciences, U.S. Department of Education (R305B070407), and from NIH (R01HD056105) to Vladimir M. Sloutsky.

References

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*, 343-382.

Clark, S. E., Gronlund, S. D., 1996. Global matching models of recognition memory: How the models match the data. Psychonomic Bulletin and Review 3 (1), 37{60.

Dennis, S., & Chapman, A. (in press). The inverse list length effect: Implications for exemplar models of recognition memory.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452–478.

Glanzer, M., & Adams, J.K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 16, 5-16.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*,16, 163-178.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: Rem-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166.

Sloutsky, V. M., & Fisher, A. V. (2004). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science*, *15*, 553 – 558.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.